Universal Sequence Replication, Reversible Polymerization and Early Functional Biopolymers: A Model for the Initiation of Prebiotic Sequence Evolution

Sara Imari Walker[1,2,3,4], Martha A. Grover[1,2] Nicholas V. Hud[1,3*]

1 NSF/NASA Center for Chemical Evolution, Georgia Institute of Technology, Atlanta, GA, USA

2 School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA, USA

3 School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA, USA

4 Current address: BEYOND: Center for Fundamental Concepts in Science, Arizona State University, Tempe, AZ, USA

Abstract

Many models for the origin of life have focused on understanding how evolution can drive the refinement of a preexisting enzyme, such as the evolution of efficient replicase activity. Here we present a model for what was, arguably, an even earlier stage of chemical evolution, when polymer sequence diversity was generated and sustained before, and during, the onset of functional selection. The model includes regular environmental cycles (*e.g.* hydration-dehydration cycles) that drive polymers between times of replication and functional activity, which coincide with times of different monomer and polymer diffusivity. Template-directed replication of informational polymers, which takes place during the dehydration stage of each cycle, is considered to be sequence-independent. New sequences are generated by spontaneous polymer formation, and all sequences compete for a finite monomer resource that is recycled via reversible polymerization. Kinetic Monte Carlo simulations demonstrate that this proposed prebiotic scenario provides a robust mechanism for the exploration of sequence space. Introduction of a polymer sequence with monomer synthetase activity illustrates that functional sequences can become established in a preexisting pool of otherwise non-functional sequences. Functional selection does not dominate system dynamics and sequence diversity remains high, permitting the emergence and spread of more than one functional sequence. It is also observed that polymers spontaneously form clusters in simulations where polymers diffuse more slowly than monomers, a feature that is reminiscent of a previous proposal that the earliest stages of life could have been defined by the collective evolution of a system-wide cooperation of polymer aggregates. Overall, the results presented demonstrate the merits of considering plausible prebiotic polymer chemistries and environments that would have allowed for the rapid turnover of monomer resources and for regularly varying monomer/polymer diffusivities.

Introduction

A key question in the origin of life is how the first biopolymers self-organized from simple chemical building blocks into increasingly complex systems with the capacity to cooperate and evolve [1]. Presently, most research that involves experimental models of early chemical evolution and function-based selection has necessarily utilized molecular systems that would have been possible only after considerable evolution of informational polymers [2–5]. Likewise, most theoretical models of early evolution have only considered how evolution could have produced the most efficient self-replicating entity after functional polymers had already emerged [6–10]. However, proposed scenarios for the chemical origins of life generally include, either explicitly or implicitly, a stage prior to functional polymer evolution when there was an initial buildup of informational polymers with random sequences – a stage that may have coincided with the selection of the first functional sequences [11, 12]. Exploring this early stage in chemical evolution is particularly difficult, being so obscured by the ensuing evolutionary history that even the identity of the first replicating material remains a subject of debate, with hypotheses ranging from RNA or some variety of proto-RNA [11, 13–16], to peptides [17], and even inorganic clay surfaces [18]. It also remains challenging to determine the environmental context from which the first informational polymers emerged, as illustrated by the wide variety of proposed sites for the origin of life, including (but not limited to): tidal shores [19, 20], deserts [21], hydrothermal vents [22], mineral surfaces [23–25], the eutectic phase of ice [26–28], deep underground [29], and even atmospheric aerosols [30].

Computer simulations offer a possible means to efficiently explore the potential merits of different scenarios for the emergence of polymer cooperativity and functionality [9, 10, 12, 31–37]. Here we explore simulations of polymer sequence and population evolution that incorporate several potentially important physical and chemical concepts of non-enzymatic replication and sequence evolution of prebiotic informational polymers, including: informational polymers with reversible backbone linkages [38–43], environmental cycling [21, 44–47], limited molecular diffusion [31, 35, 36], monomer recycling [16, 48–50], and template-directed synthesis without sequence restrictions [51]. Here we introduce the term Universal Sequence Replication (USR) to represent the possibility that prebiotic template-directed synthesis provided a means for the replication of polymers of a particular chemical structure (*i.e.* backbone and side chain structure), regardless of monomer sequence (such that replicative rate constants are, at least to first-order approximation, sequence-independent). USR should be distinguished from the more general framework of non-enzymatic template directed synthesis, which includes chemistries where certain sequences may have a replicative advantage; USR can be seen as a special case where the intrinsic fitness landscape is flat and no sequences possess an intrinsic replicative advantage.

Although the framework presented here shares some individual features with other models, their unification in the model presented provides a fresh perspective on a plausible scenario for environmentally-driven early emergence and evolution of informational polymers. Moreover, we diverge from most previous models in that we address the likely possibility that a nascent population of prebiotic informational polymers would have initially possessed no functional sequences (including no replicase activity). As such, we draw a distinction between the autonomous or mutually dependent "replicators" that have been explored in many previous models (*e.g.* quasispecies and hypercycle models [6–8]), and the implications of environmentally-driven USR that we explore here. This distinction highlights that the former class of models relies upon the self-replicating capacity of a particular polymer sequence or group of sequences, *i.e.* these models assume that functional polymers are already present in the extant population; whereas in the model explored here such a complex function has not yet emerged. Instead, USR posits that all polymers, regardless of monomer sequence, were on equal footing prior to the emergence of functionality.

Kinetic Monte Carlo simulations were used to explore the dynamics of populations of informational polymers that are formed by spontaneous polymerization, replicated by USR, and subject to hydrolysis in a diffusion-limited environment. The simulations reveal that a population of nonfunctional polymers is evolvable in the sense that new polymer sequences are continually introduced to a diverse extant population, and selection for new functional sequences can occur. The observed dynamics provide insights into a robust mechanism for the early exploration of sequence space, including how functional sequences might become established in an initially random sequence pool. Although USR imposes a flat replication fitness landscape, local feedback resulting from localized resource recycling and limited-diffusivity permit selection of functional sequences. A key result of our simulations is that functional selection does not dominate the system dynamics and species diversity remains high, where we define a species as a population of polymers with identical sequence. High species diversity during functional evolution permits the emergence and spread of more than one functional sequence, where nucleation of functional sequences may be temporally and/or spatially separated. Furthermore, functional sequences need not become permanently fixed within a population to beneficially impact population level dynamics. This rudimentary form of polymer cooperativity illustrates how evolution may have progressed at a level of the polymer pool before enzyme-based polymer replication or compartmentalization had emerged. The results presented suggest chemical features of candidate polymers and environments that might have initiated functional evolution of informational polymer populations in the origin of life.

Methods

In constructing our model, we implemented a relatively small number of parameters to explore select chemical features of candidate prebiotic polymers and their physical setting, as well as the feedback between polymer and monomer populations. We first qualitatively outline the most salient features of the prebiotic scenario captured by our model, and then describe the specific details of how the model is numerically implemented.

Model Description

In this section, we briefly describe the physicochemical features of our model, along with the associated adjustable model parameters. All model parameters (apart from enzymatic functional activity) are sequence-independent and are quoted in dimensionless units (see Supporting Information for a detailed description of model parameterization, Methods S1, including a table of parameters and values used in simulations, Table S1). Our primary motivation in choosing *all* kinetic rates to be sequence-independent was to focus on the role of the environment (*e.g.* cycling, diffusion) in driving system dynamics and functional selection (without biases introduced by differential kinetic rates). One exception is enzymatic functional activity, which is sequence-dependent in our model.

Regular environmental cycles. Environmental cycles would have occurred regularly on the prebiotic Earth (*e.g.* day-night, tidal, seasonal, hot-cold, freeze-thaw, hydration-dehydration, etc.). In the model presented here we appeal to hydration-dehydration cycles, such as those driven by tidal fluxes or day-night cycling, to provide an energetic source for the assembly of monomers into polymers, where physicochemical properties vary with the environmental phase. Polymerization via spontaneous assembly and USR via template-directed synthesis occur during the hot-dry conditions of the dehydrated phase. Polymer degradation and diffusion of monomers and polymers occur in the cool-wet conditions of the hydrated phase. Additionally, functional polymers (when present in the extant population) only exhibit catalytic activity during the hydrated phase, when cool-wet conditions promote the folding of a polymer into its active state.

Reversible polymerization. The model is based upon informational polymers with reversible backbone linkages [38–43]. During the hydrated phase, condensation polymers are subject to spontaneous degradation (hydrolysis), governed by the first-order rate constant k_h. Monomers liberated via polymer hydrolysis are added back to the local monomer population, creating localized feedback between polymer and monomer concentrations.

Finite monomer concentration. The total number of monomeric units (*e.g.* nucleotides) in monomer and as polymeric residues is constant in our model system, with equal numbers of the two monomer species labeled A and B. The choice to focus on closed mass systems was motivated by our goal of connecting to chemically realistic scenarios, where monomer concentration would have been naturally limited by available resources. For systems with a finite supply of resources, sustainable polymerization requires recycling through turnover of resources via polymer hydrolysis [16, 48–50]. An exception is the case where new monomers are generated by a functional polymer that acts as a monomer synthetase, where monomer production is still limited by the availability of precursor molecules, which are also a finite resource (see below).

Surface confinement and limited diffusion. Incomplete mixing, limited diffusion, and surface attachment have been shown to promote template-directed synthesis [51, 52] and to limit the deleterious effects of "parasites" (non-functional polymers) on selection of functional sequences [31, 35, 53]. In the present model, we therefore confine polymer and monomer movement to a surface, as might occur on a mineral with a thin film of water (or a more viscous solution) covering the surface, or other surface-binding phenomenon that permits the reversible association of monomers and polymers and limited movement in two dimensions. The reaction-surface is modeled as a square lattice (see below) where diffusion of monomers and polymers between neighboring sites is governed by the hopping rates k_m and k_p, respectively, subject to the physical constraint condition $k_m \geq k_p$. Diffusion only occurs during the hydrated phase of the cycle, when water activity is high. During the dehydrated phase, when the surface is dry, or the viscosity of the thin film is high, diffusion between lattice sites is considered to be negligible, with $k_m = k_p = 0$.

Universal sequence replication (USR). Recent studies have shown that altering the chemistry of nucleic acid backbone linkages (*e.g.* with a reversible linkage) [38,39,41,43], or binding of template strands to a surface (*e.g.* reduced mobility compared to monomers) [51], allows for accurate template-directed synthesis for a wide range of sequences (*i.e.* approaching a form of USR). The effects of idealized USR are explored in our model, *i.e.* all polymers have the same intrinsic rate constant for replication, parameterized by the third-order rate constant k_r. Despite the fact that all polymers share the same inherent replicative fitness, replication propensity is not the same for each polymer in each replication cycle, as the replication probability for a given polymer depends on both k_r and the local monomer concentrations. Stochastic variation in the spatial distribution of free monomers results in spatiotemporally varying rates of polymer replication, creating a dynamic fitness landscape. As shown below, in the absence of functionality, this selective advantage is random, acting on local populations rather than individuals, with populations sequestering the most resources having an increased chance of survival, independent of their sequence distribution.

Spontaneous polymer assembly and exploration of sequence space. To simplify our model, we invoke spontaneous polymer assembly as the only means to generate novel sequences (*i.e.* we do not consider the effects of mutations during replication or of genetic recombination, which have been extensively studied elsewhere, *e.g.* see [6, 8, 31, 35, 37, 54, 55]). In this framework, polymer degradation allows continual exploration of sequence space as novel sequences are introduced via spontaneous assembly of free monomers. The rate of spontaneous assembly (to be contrasted with template-directed assembly) is governed by the second-order rate constant k_s, and local monomer concentrations. Throughout this work we set $k_s = 10^{-7}$, corresponding to a maximum global production rate of approximately 1.5 new sequences per dehydrated phase when all monomers are free (with lower rates when some monomers are sequestered in polymers). This value is intended to reflect a less efficient process of spontaneous polymerization as compared to template-directed synthesis.

Emergence of a functional polymer. A critical stage in the origin of life was the emergence of the first functional informational polymers [56]. We therefore explore with our model the case where a functional sequence is discovered by the random appearance of the sequence. Ma and coworkers have previously argued that in an RNA world, where polymer replication is accomplished without the need for enzymatic polymers (*i.e.* by some mechanism of USR [51, 57, 58]), that a nucleotide synthetase was the first enzyme to emerge [34]. A synthetase would have a clear advantage in an environment where local monomer concentration is the limiting factor for replication. We therefore explore a similar test case. In our simulations, the first polymer to emerge with a beneficial function is referred to as an *A*zyme, a polymer capable of synthesizing *A* monomers from an additional but also finite resource, proto-*A* (*pA*). The catalytic activity of the *A*zyme is governed by the second-order rate constant k_A and local *pA* concentration. The Azyme thereby directly couples the functional dynamics to the local environment (*e.g.* local monomer abundance). Additionally, the emergence of a synthetase captures some properties of metabolic replicator models [34, 37], demonstrating a possible pathway between the pre-functional stage of replicating nonfunctional sequences and the subsequent stage of functional sequence optimization.

Kinetic Monte Carlo Implementation

The simulations were implemented via a spatially-explicit hybrid kinetic Monte Carlo algorithm [59, 60]. Here we provide a brief summary of the technical details of our implementation. A more in depth discussion of the numerical implementation is provided in the Supporting Information (Methods S1).

Initial conditions and the reaction surface. The reaction surface is modeled on a 64 × 64 square lattice. Each lattice site represents a locally homogenous reaction domain, characterized by a freely interacting community of monomers and polymers (*i.e.* a locally well-mixed environment). The lattice size of 4096 sites was chosen to be small enough for numerical tractability, yet large enough to allow spatial correlations to be observed for the range of kinetic and diffusive parameters under study (such that any spatial organization observed for the parameter ranges explored here is smaller than the lattice dimensions). We impose periodic boundaries to avoid edge effects. The total number of monomeric units (*e.g.* nucleotides) in monomer and as polymeric residues is constant, with equal numbers of the two monomer varieties labeled A and B. The initial conditions are homogeneous, with all mass in monomer: each of the 4096 lattice sites is initialized with 60 A and 60 B monomers, such that $245,760$ monomers each of species A and B are evenly distributed over the reaction domain at the start of a simulation run. For functional runs in which the Azyme appears, each site is also initialized with 60 pA monomers, and with 60 pB monomers for simulations where the Bzyme also appears. No polymers are present at the start of a given simulation run (an exception is for functional runs, which start from an already established pool of random sequence polymers at quasi steady-state, see below).

Defining the polymer species pool. The number of possible polymer species, N, increases exponentially with polymer length: for a polymer of length L and two residue types there are $N = 2^L$ possible sequences. We chose to focus our study on polymer populations with a fixed length R_L, meant to approximate the dynamics of a population with a mean length of $\bar{L} = R_L$ residues. For the simulations presented here, $R_L = 20$. Our studies of the dynamics with other fixed polymer lengths, including $R_L = 2, 6$ and 10, yielded qualitatively similar results. Although using shorter length sequences would have enhanced numerical efficiency, we used oligomers of length 20, as nucleic acids of this length are sufficiently long to adopt folded structures and, arguably, large enough to exhibit initial levels of catalytic activity [61]. Additionally, for $R_L = 20$ the sequence space is sufficiently vast that our simulations never explore all possible sequences within the timescales of our simulations. In other words, the sequence space is larger than what our system can dynamically explore in the space and timescales under study; satisfying the minimal requirement of a system with the potential for unlimited heredity and thus open-ended evolvability [1].

Assigning polymer composition. Our initial model implementation focused on polymers with fixed length R_L and any possible sequence diversity (*i.e.* any possible arrangement of A and B monomer residues could appear). However, sequences with a ratio $A/B = 1$ (*e.g.* composed of 10 A monomers and 10 B monomers for $R_L = 20$) were an attractor for the dynamics under study, having the greatest access to available resources. Any deviations from the mean distribution, yielding a sequence population dominated by sequences with $A/B \neq 1$, quickly returned to populations dominated by $A/B = 1$ (as an example consider a pure A-residue sequence which only has access to half the resources in the reactor pool and would quickly be outcompeted by other sequences containing some B residues which had greater resource availability). Therefore, the simulations were set such that every sequence nucleated contained both 10 A

monomers and 10 B monomers. This greatly simplified polymer identification since each unique species need only be identified by a single ID or lineage number, i (a dramatic simplification over tracking the specific sequence structure of each unique lineage). This approximation reduced the size of the relevant sequence space from $N = 2^{20} = 1,048,576$ to $184,756$ possibilities, but still retained our requirement that the relevant sequence space be much greater than what our dynamics could spatiotemporally explore within a given simulation run (see previous section), and allowed larger and longer statistical samplings.

Coupling of diffusive and kinetic events to environmental cycles. Our numerical implementation of the processes outlined in the previous section explicitly takes into account environmental cycling in order to accurately capture the dynamics of populations of condensation polymers with reversible linkages. The kinetic Monte Carlo implementation is therefore partitioned into two phases: a dehydrated phase, where all lattice sites are diffusively isolated (*i.e.* diffusion is turned off); and a hydrated phase, where lattice sites interact diffusively through polymer hopping events and monomer diffusion. Polymer assembly and replication occur only in the dehydrated phase, and polymer degradation occurs only in the hydrated phase. For simulations exploring the emergence of a functional sequence in the extant population, an additional kinetic process describing enzymatic catalysis is included in the hydrated phase.

The dehydrated phase. During the dehydrated phase, each lattice site x on the two-dimensional lattice is diffusively isolated and treated individually with the standard Gillespie algorithm [62]. The dehydrated phase supports the kinetic processes of spontaneous assembly and replication. The probabilities of reaction events are weighted by their relative reaction propensities. Spontaneous polymer assembly occurs with (second-order) reaction propensity

$$a_s \quad = \quad k_s AB \ , \tag{1}$$

and USR via template-directed assembly occurs with (third-order) reaction propensity

$$a_{r_i} \quad = \quad k_r AB \ N_i \ . \tag{2}$$

Here A, B, and N_i are the number of individual A monomers, B monomers, and polymers of lineage i (the species ID number) at a specific site x. The total rate for polymer assembly and USR via template-directed assembly are governed by their respective rate constants, k_s and k_r, *and* the amount of available monomer resource. This dependence is essential to study how local resource availability affects the dynamics of polymer populations (for example, this feature leads to nontrivial spatial patterning, see Results). Since system dynamics are environmentally driven, a polymer may copy itself *at most* once per hydration/dehydration cycle. Therefore, after each template-directed replication event the total number of polymer templates available for further synthesis is decreased by one unit. In other words, we explicitly take into account that the template will not dissociate from the substrate until the next hydrated phase when dilution can drive dissociation. We note that cycling is often invoked as a mechanism for driving strand dissociation [63], however, it is not always explicitly included in model dynamics. We explicitly include cycling, along with its impacts on generational turnover, to explore the potential evolutionary impact on our model prebiotic system.

We note that USR via template-directed assembly is treated as a third-order process with a total rate dependent on resource availability through the nucleation term AB (in addition to the rate constant k_r and availability of the template N_i - see eq. 2). This relationship is meant to treat dimer formation as the rate limiting step for formation of full-length polymers, *i.e.* for the chemistries under investigation here, dimer association with a template is much stronger than monomer association. USR is treated as a one-step process for formation of full-length sequences, since the majority of polymers will form quickly once the first step of dimerization occurs. We therefore consider these approximations to be consistent with the physicochemical processes modeled, with the benefit that they lead to a simplified implementation of the numerical algorithm while still permitting resource dependence in the rates. The choice of kinetics for spontaneous polymer assembly has similar motivation. Dimers are expected to have stronger association to the reaction surface - *e.g.* mineral or clay - than monomers, thereby promoting polymerization once dimerization has occurred. The choice to implement resource dependence via a nucleation term AB (rather than AA or BB) is consistent with our implementation of a reduced sequence space.

The hydrated phase. During the hydrated phase, monomers and polymers diffuse, and polymers may degrade or, if functionally active, perform catalysis. Individual lattice sites are diffusively coupled in the hydrated phase, and the dynamics are therefore modeled with a spatially-explicit hybrid kinetic Monte Carlo algorithm [59, 60]. All kinetic events are treated locally, occurring within an individual lattice site which is modeled as a locally homogeneous reaction domain, and only diffusive events occur between sites. A natural partition between rare and common events occurs due to the large separation in the population densities of monomer and polymer observed in our simulations. The simulations are therefore hybridized such that monomer site hopping is coarse-grained and treated via mass-action kinetics (see Supporting Information for more details, Methods S1). All other events in the hydrated phase are treated stochastically. We have verified that a full stochastic treatment (via a standard spatially-explicit Gillespie algorithm [64]) reproduces our hybrid results. Against the background of coarse-grained monomer diffusion, the rare polymer events of diffusion and degradation occur. The probabilities of (rare) reaction events are weighted by their relative reaction propensities. Polymer hydrolysis (degradation) occurs with reaction propensity

$$a_{h_i} = k_h N_i \ , \tag{3}$$

and polymer diffusion (hopping between nearest-neighbor lattice sites) occurs with propensity

$$a_{p_i} = k_p N_i \ . \tag{4}$$

Here hydrolysis is all or none, and degradation is therefore treated as a first-order one-step process which is stochastically determined. This approximation is made based on the implicit assumption that shorter polymer lengths are less stable, as might occur for cases where 20mers can maintain stable folded conformations whereas shorter length polymers cannot (*i.e.* we assume increasingly shorter length polymers become increasingly less stable to hydrolysis).

Data analysis. Data presented for quasi-steady state distribution averages, for explorations of both kinetic (*i.e.* replication/hydrolysis) and diffusive parameter space, are the combined result of time-averages over 2500 cycles and ensemble averages over a small statistical sampling of runs (averaged over 5 and 10 runs for kinetic and diffusive parameter space exploration, respectively). Small statistical samples are sufficient given the small spread in simulation values and the length of simulations with large time-sampling statistics. The quasi-steady state distribution is defined as the period when the ratio of polymer to monomer achieves an equilibrium value (with fluctuations due to stochastic effects). We used the term “quasi” here to indicate that the sequence population is not static, even at steady-state (see Results). Quasi-steady state distributions are calculated starting at $t = 2500$ cycles (typically steady-state is achieved at $t = 500 - 1000$ cycles depending on simulation parameters). Time averages were taken from $t = 2500 - 5000$ cycles. Data point error bars correspond to sample standard deviation on the mean time-averaged values. In comparing kinetic and diffusive processes for the results presented in this work it is important to note that the simulation dynamics are dependent on the overall rates of processes, which are dependent on both monomer and polymer abundances. The kinetic rate constants (k_s, k_r, and k_h) and the diffusive hopping rate constants (k_m and k_p) are useful in providing measures of the relative strengths of the processes under study, but must not be confused with the actual rates for the different kinetic and diffusive processes, which are dynamically determined by the ratios of monomer to polymer in a given simulation and their spatial distribution (as defined above, eqs. 1, 2, 3, and 4).

Simulating functionality. In simulations including the Azyme, which catalyzes formation of A monomers from pA monomers, two additional processes are added to the hydrated phase, diffusion of pA monomers and catalytic conversion of $pA \to A$. Diffusion of pA monomers, like diffusion of A and B monomers, is treated via mass-action kinetics in the hybridized algorithm. Enzymatic catalysis by the Azyme is added to the rare events of the hydrated phase, with the probability of catalysis calculated from the reaction propensity

$$a_{pA} = k_c \; pA \; N_{\mathrm{Azyme}} \; , \tag{5}$$

where pA is the number of pA monomers on a local site x, k_c is the catalytic rate constant for conversion of $pA \to A$ in the presence of the Azyme, and N_{Azyme} is the total number of Azymes on the local site x. To illustrate the impact of the emergence of a functional sequence, data was saved at $t = 2500$ cycles for all details of a given simulation run. This data provided the initial starting distribution of monomers and polymer species for the functional runs. The choice of starting at $t = 2500$ cycles is somewhat arbitrary given that the systemic dynamics in the quasi-steady state evolution are time-independent, but was chosen to be sufficiently late in the system evolution that a quasi-steady state had been established (*i.e.* the ratio of monomer to polymer was relatively constant). To this initial condition, 60 pA monomers were added to each site (to model a previously untapped resource in the environment) and a single polymer representing the Azyme sequence, or a nonfunctional sequence, was inserted on the lattice as a spontaneous assembly event. Results were averaged over twenty-five runs, each with a randomly chosen insertion point for the inoculated sequence. Selection of the inoculation site was weighted by the propensities for spontaneous assembly (*i.e.* inoculation was not completely random but determined by the resource distribution in the system as done for any other spontaneous assembly event). The simulations were permitted to run until the inoculated sequence (functional or nonfunctional) died out, or until the sequence had survived for 5000 cycles. Lifetimes were averaged over survival times for the inoculated sequence lineage (defined as the duration of time where at least one individual of the inoculated sequence is still on the lattice) taken over twenty-five simulation runs. Population size averages, the average number of extant species, and exploration rate were averaged over the sequence lifetime. For example, a polymer that lives 5 cycles is only averaged over 5 cycles, and therefore yields much higher variance in the data than nonfunctional simulations, which are averaged over entire populations of thousands of sequences, over thousands of cycles. Simulations including a Bzyme, catalyzing $pB \to B$, were inoculated in a similar manner with the initial simulation time taken at $t = 4000$ cycles.

Results

For each simulation run, the model system was initialized with 60 A and 60 B monomers at each of the 4096 lattice sites, and no polymers. Starting from this homogeneous distribution, we tracked the spontaneous assembly, replication and spatial propagation of informational polymers over several thousand hydration-dehydration cycles. All stochastic events occur locally (*i.e.* within or between neighboring lattice sites); however, a global dynamic between local communities emerges due to diffusive contact. Stochasticity is observed to drive polymer population dynamics. In particular, for the parameter ranges investigated here, no system ever achieved a stationary steady-state population of polymers *with fixed sequence information*. Instead, the sequences represented in the polymer population continually change with time (Figure 1A). The *total population* of polymers is maintained once the system reaches equilibrium (Figure 1B), but the population is temporally varying with respect to the distribution of polymers among existing sequences and due to the appearance of new sequences – a state of dynamic kinetic stability (DKS) [65,66]. Specifically, as individual polymers degrade, monomer recycling allows a quasi-steady state number of polymers to change in sequence distribution by providing resources for replication and spontaneous polymerization. Moreover, stochastic fluctuations in the number of individuals with a given sequence cause species to have a finite lifetime, while some new sequences that appear even after DKS has been achieved are observed to take hold and propagate in the population (Figure 1). We have verified that the quasi-steady states of the DKS observed in our simulations subsist for thousands of environmental cycles once an equilibrium distribution between monomer and polymer is established (running simulations upwards of $> 20,000$ cycles). Stochasticity is also observed to drive dynamic pattern formation in the spatial distribution of polymers. A common feature of our simulations, for a wide range of parameter space, is the spontaneous appearance of polymer clusters. These localized high concentrations of polymers typically begin as a large number of small clusters that then coalesce into fewer, larger clusters over time (Figure 1C; time evolution movies of cluster formation are provided in Supporting Information, Movie S1 and S2). Depending on the parameters of a given simulation, the space between clusters can be essentially devoid of polymers. An individual cluster is typically composed of multiple polymer sequences that mutually benefit from being part of a cluster. As will be shown below, cluster formation correlates with several important system characteristics, such as local monomer concentration, polymer lifetime, sequence diversity and functional sequence propagation.

Exploring Replication and Hydrolysis Rate Parameter Space

To explore system characteristics as a function of specific model parameters, we measured and compared ensemble averaged data for a number of system metrics collected for simulation runs in which one or two parameters were varied, with the remaining model parameters held constant. For the first set of simulations presented, the polymer replication and hydrolysis rate constants k_r and k_h, respectively, were varied while the rate constants for spontaneous polymer assembly and for monomer and polymer diffusion, k_s, k_m and k_p, respectively, were held at fixed values. Since our aim is to investigate systemic features and evolutionary potential prior to the onset of functionality, no functional sequences exist in the systems presented in this section. As demonstrated by the data shown in Figure 2, the six system metrics of Average Sequence Lifetime, Average Species Population Size, Number of Extant Species, Total Polymer Population, Sequence Exploration Rate, and Average Local Diversity, each change to varying degrees in response to different values for k_r and k_h.

Of the six metrics shown in Figure 2, the dependence of Average Species Lifetime on variations in k_r and k_h is, perhaps, most intuitive. A species is defined as a population of polymers sharing the same unique sequence of A and B monomers, and the species lifetime is the number of contiguous cycles in which one or more copies of that sequence exists in the system. Due to the large number of possible sequences, it is assumed that a particular sequence will spontaneously appear in the system at most once over the timescales of interest. Figure 2A shows that Average Species Lifetime increases with replication rate constant k_r, since higher polymer replication rates result in more *copies* of an extant sequence existing in the system. Species lifetime decreases with increases in the hydrolysis rate constant k_h, since all polymers have an increasing probability of spontaneous degradation. For the case of $k_h = 1$, a polymer has a 63% chance of degradation during each hydrated phase. Thus, most polymers do not survive into the next cycle, when they would have the opportunity to replicate. Consequently, for $k_h = 1$ average polymer lifetime is < 1 cycle for all values of k_r considered. In the case of no replication, when $k_r = 0$, only one individual of any given sequence will ever exist in a simulation and therefore the Average Species Lifetime is determined only by the rate of polymer hydrolysis (*i.e.* by the average lifetime of a single polymer).

For $k_r \neq 0$ and $k_h < 1$, we observe that the ratio of k_h/k_r largely determines Average Species Lifetime (note: the ratio k_h/k_r represents the ratio of *rate constants* and not the ratio of effective reaction rates which are dependent on local monomer and polymer densities). For $k_h/k_r = 10$, a lifetime $> 10,000$ cycles is observed, while for $k_h/k_r = 1000$, an average lifetime in the range of $100-1000$ cycles is observed. For values of $k_h/k_r \geq 10,000$, the Average Species Lifetime drops below 10 cycles. A distinction can be made in Figure 2A between systems with an Average Species Lifetime < 10 cycles and systems with an Average Species Lifetime of > 100 cycles. In the former, sequence lineage (*i.e.* species) propagation is minimal, whereas in the latter case, propagation of sequence lineages is robust. This result is also illustrated by other system metrics presented below.

Figure 2B shows the Average Species Population Size for extant species after DKS has been reached. Two regimes are clearly visible. In simulations with $k_h/k_r \geq 10,000$, the Average Species Population Size is < 2. As with species lifetime, a high hydrolysis rate will limit the ability for polymers to replicate even for large values of k_r, thereby limiting the propagation and copy number of any given species. For values of $k_h/k_r < 100$ (with $k_r > 0$), Average Species Population Size reaches a plateau value that is positively correlated with k_r. When $k_r = 0$, the case of no polymer replication, there can only be one copy of each sequence in the system regardless of hydrolysis rate, as is observed in Figure 2B.

Figure 2C provides a view of how the Number of Extant Species, defined as the average number of unique species present at any time after DKS has been reached, varies with k_r and k_h. Again, for a sufficiently high hydrolysis rate (*i.e.* $k_h = 1$), polymers generated in a given cycle are likely to degrade before having the opportunity to replicate, thus limiting the total number of polymers present in the system and thereby the Number of Extant Species. For $k_h/k_r \lesssim 1000$, the Number of Extant Species reaches plateau values that decrease with increasing replication rate constant. This relationship is a direct result of competition for a finite supply of resources: larger replication rates lead to greater competition for the limited supply of resources, resulting in higher species extinction rates. In the case of no replication (*i.e.* $k_r = 0$), the Number of Extant Species is equal to the Total Polymer Population Size, *i.e.* the steady-state value is determined only by k_h, when all other parameters are held constant.

The Total Population Size of a system, shown in Figure 2D, is defined as the total number of polymers (regardless of sequence) present after DKS has been reached. Plots of this metric illustrate that for $k_h/k_r \leq 100$, a plateau value of approximately $21,000$ total polymers is reached. This value corresponds to roughly 85% of monomers being sequestered in polymers, and 15% as free monomers. This upper limit on the number of monomers that can be incorporated into polymers is the result of an artificially imposed limitation on the simulation dynamics. Specifically, replication or spontaneous polymer formation cannot take place within a local environment that contains less than the number of monomers necessary to make a full-length polymer (*i.e.* twenty monomers, in our simulations with polymer lengths fixed at twenty). The combined systemic features shown in Figures 2B, 2C, and 2D show that for a wide range of hydrolysis and replication rates the same total number of polymers will be present in the system, but the population will be divided between a smaller number of unique sequences (or extant species) as the rate of replication is increased.

The metric of Sequence Exploration Rate is defined as the rate at which new sequences appear in a given system. Because new sequences arise solely by spontaneous formation, this metric provides a measure of the systemic ability to explore sequence space, an absolute necessity if a system is to evolve through the spontaneous appearance of polymers with functional activity. As shown in Figure 2E, for systems where k_h and k_r values give rise to a considerable percentage of free monomers in a state of DKS (*i.e.* small total polymer population sizes), the rate of sequence exploration is only limited by the parameter k_s, the rate constant for spontaneous polymer formation. As hydrolysis rates are decreased and replication rates increased, the number of free monomers available for spontaneous polymer formation decreases, thereby decreasing the Sequence Exploration Rate. For the case of $k_r = 0$, where polymers are generated only by spontaneous assembly, Sequence Exploration Rate remains high for a wide range of k_h values, being only limited by the number of monomers made available via polymer hydrolysis. However, the special case of $k_r = 0$ also corresponds to a system in which no sequences propagate through replication. Thus, for a system to evolve through the spontaneous discovery *and* propagation of a functional sequence, it is necessary that both the Sequence Exploration Rate be nonzero and that the Average Species Population Size be greater than one. For the system parameters explored here, values of k_h/k_r between 100 and 1000 appear to be within a "sweet spot" of compromise between sequence exploration rates and the ability for a sequence to take hold in a system through replication.

Average Local Diversity is the final system-level metric shown in Figure 2. In contrast to the previous five metrics, Figure 2F describes the *spatial distribution* of sequences in the system. Sequence diversity was quantified using the Shannon entropy equation [67], and was calculated locally at each site on the lattice and averaged over the total number of lattice sites (see Supporting Information for additional mathematical details, Methods S1). Average Local Diversity therefore provides a statistical measure of the extent to which multiple sequences coexist on the same lattice site. As such, it provides a measure of diffusive mixing of populations (discussed below) and competition, whereby spatial regions with low local diversity result either from low diffusivity or high rates of local resource competition that result in fewer unique species. Unlike the previous plots, Average Local Diversity is strongly dependent on k_h, rather than the ratio k_h/k_r. Low hydrolysis rates promote high local diversity, since extinction rates are low with high Average Species Lifetimes.

Exploring Diffusive Parameter Space

We now present simulations designed to explore the effects of monomer and polymer diffusion rates on the system metrics defined above. Specifically, the polymer diffusive hopping rate, k_p, was varied within the range $0 \leq k_p \leq 1.0$, and the monomer diffusive hopping rate, k_m, was varied within the range $0 \leq k_m \leq 90$. The simulation values investigated are subject to the condition $k_m \geq k_p$, since polymers cannot diffuse faster than monomers. For the simulations presented, the remaining system parameters (*e.g.* the kinetic rate constants) were held fixed at values corresponding to intermediate metrics observed for the simulations presented in Figure 2 (*i.e.* $k_r = 10^{-4}$; $k_h = 0.1$; $k_s = 10^{-7}$). As in the previous section, our aim in this section is to investigate systemic features and evolutionary potential prior to the onset of functionality. Therefore, no functional sequences are present. Thus, all observed system characteristics presented in this section (including the onset of spatial patterning) arise in the absence of any polymeric catalytic activity (*i.e.* without functional sequences).

Spatial patterning. Before discussing system metric results, it is instructive to consider the effects of varying k_m and k_p on the spatial distribution of polymers, as these diffusion-dependent distributions are important for understanding results for all other system metrics, as well as being interesting in their own right. As shown in Figure 3, simulations that have completed 3000 hydration-dehydration cycles exhibit polymer spatial organization that is highly dependent on the diffusivities k_m and k_p, and their relative magnitude. All systems shown in Figure 3 had identical initial conditions – the system was initialized with a uniform distribution of monomers and no polymers at $t = 0$. For monomer diffusive hopping rates $k_m \geq 0.1$, the total number of polymers in the system is fairly constant, with approximately $10,000$ polymers on the lattice. However, large variations in the distribution of polymers are clearly visible. In simulations where monomers and polymers have low diffusivities (*e.g.* $k_m = 0.01$ with $k_p = 0.001$), polymer species tend to stay spatially isolated and localized near their nucleation sites. Localized resource recycling sustains these small communities, with small polymer clusters being homogeneously distributed across the simulation lattice. As k_m is increased from 0.01 to 10, with k_p fixed at 0.001, polymer cluster size gradually increases until only one or two dominant clusters are observed after 3000 hydration-dehydration cycles. A similar trend of increasing cluster size is observed for simulations with a 10-fold greater polymer diffusion rate, (*e.g.* $k_p = 0.01$), with larger, more diffuse clusters observed for greater polymer diffusion rates. Further increase in k_p (*i.e.* $k_p = 0.1$ and 1.0 in Figure 3) leads to a loss of defined clusters, or clustering on a scale that is larger than the simulation lattice. In simulations with no polymer movement, $k_p = 0$, new sequences can only be introduced at a grid point by spontaneous polymerization, and clustering is not observed regardless of k_m value. Additional spatial maps are provided in Supporting Information (Figures S1-S5).

The clustering patterns observed in Figure 3 emerge due to the underlying stochasticity of the dynamics. As the first polymers nucleate and replicate, stochastic fluctuations in populations lead to inhomogeneities in polymer population densities. Populations with a slight initial excess of polymer grow by sequestering free monomers that diffuse to their local vicinity. Isolated polymers migrate toward these regions of concentrated resources or go extinct due to resource competition. Thus, clustering emerges as a result of an indirect form of cooperativity between replicating polymers, where early populations with higher polymer densities gain a fitness advantage. Polymers that exist within a cluster enjoy a local recycling dynamic in which the polymers of a cluster act as a reservoir of monomers: fresh monomer resources for replication become available upon polymer hydrolysis, where monomers are sequestered into polymers before these recycled resources have the opportunity to diffuse away from the cluster. This dynamic can occur because the overall rate of polymerization within a cluster is larger than the polymer diffusion rate, thereby resulting in localization of polymer populations (for polymer diffusivities with $k_p \geq 0.01$ in Figure 3 strong clustering is not observed). Between the clusters, polymer density fluctuates near zero. These low population regions act as physical barriers to the transport of information between clusters, leading each aggregate to have a unique sequence population. In contrast, these polymer-depleted regions permit monomer transport, which, through stochastic fluctuations, allow growing clusters to acquire resources from shrinking clusters, even though the polymer clusters may not be in direct contact. The results shown in Figure 3 also illustrate that, for the parameters used in these simulations, polymer population growth is greatly inhibited if monomer diffusion is too slow. In particular, simulations carried out with $k_m = 0.001$ were almost devoid of polymers.

Diffusive dependence of system metrics. In Figure 4 the six system metrics are shown for the set of simulations in which k_m was varied from 0.001 to 90, and k_p from 0 to 1.0. For nonzero polymer diffusion rates, Average Species Lifetime tends to decrease with monomer diffusion rates of $k_m \geq 0.1$ (Figure 4A). This trend is associated with the observed clustering at higher monomer diffusion rates. High diffusion rates allow for a small number of species to spread and sequester the majority of available resources, which causes high extinction rates for later-nucleating sequences. Thus, a species will, *on average*, have a shorter mean lifetime when a smaller number of species are able to dominate the sequestration of resources. The effect of diffusion on species growth is perhaps more easily appreciated by considering two other system metrics: Average Species Population Size and Number of Extant Species. Average Species Population Size (Figure 4B) shows a positive correlation between population size and polymer (and monomer) diffusion rates, with the average population size increasing with increased diffusivity. This result illustrates that increasing polymer and monomer diffusion rates leads to a decrease in the number of species that are able to dominate the acquisition of resources. Likewise, the Number of Extant Species (Figure 4C) shows that the mean number of extant species, like Average Species Lifetime, decreases with increasing polymer and monomer diffusion rates. In the special case of no polymer diffusion ($k_p = 0$), Average Species Lifetime and Number of Extant species increase, for the most part, with increasing monomer diffusion rates. This distinct trend is apparently due to increased monomer diffusion rates allowing for the replication of polymers before spontaneous hydrolysis, but without the ability for sequences to spread and "colonize" other regions of the surface, which also limits species population size. As noted above, the total number of polymers in a simulation after 3000 cycles was relatively independent of k_p, for $k_m \geq 0.05$ (Figure 4D).

In contrast to the results presented above for variations in hydrolysis and replication rates, we observe that the Sequence Exploration Rate (Figure 4E) is essentially independent of polymer diffusion rate and only modestly dependent on the monomer diffusion rate for the parameter ranges explored in Figures 3 and 4. Thus, when considering the ability for a system to evolve through sequence exploration, variations in monomer and polymer diffusivities are more likely to affect the capacity for survival of a newly nucleated sequence than the system's capacity to discover new sequences. For example, in the case of no polymer diffusion ($k_p = 0$), a sequence is not able to reap the benefits of spatial expansion, which permits population growth. On the other hand, a sequence that emerges in a system with high polymer mobility will experience strong competition, and may not take hold in the system. The ability for a *functional* sequence to overcome these pressures in explored below. Finally, the metric Average Local Diversity shows a unique response to changes in monomer and polymer diffusion rates. Like Average Species Population Size, Average Local Diversity increases with k_p. This positive correlation with polymer diffusion illustrates how more rapid polymer movement leads to more complete spatial mixing of polymer species. In contrast, unlike the five other system metrics, Average Local Diversity is apparently independent of k_m, an observation that, taken with other system observations, demonstrates how variations in monomer diffusion rates can affect the spatial distribution of polymers without affecting the local spatial distribution of species diversity.

Demonstrating the Emergence of Functionality

We now address the potential for an individual, catalytically active sequence to become established in a pre-existing pool of nonfunctional polymers. As introduced in The Model section, we chose for our test case the emergence of a polymer sequence (the Azyme) that catalyzes the production of A monomers from a previously untapped resource of proto-A (pA) monomers. For these simulations, the system was initialized with 60 pA monomers at each lattice site (in addition to the 60 A and B monomers), with a single Azyme sequence being introduced after the system of nonfunctional polymers had reached a state of DKS. For the results presented here, the Azyme was introduced at $t_A = 2500$ cycles. The catalytic rate constant of the Azyme was set sufficiently high that enzymatic activity would only be limited by access to pA monomers. This criterion was satisfied by setting the catalytic rate of the Azyme such that one enzyme would be able to convert all pA monomers within its lattice site to A monomers within a single hydrated phase. Thus, the observed impact of the Azyme sequence on the system does not depend on the catalytic activity of the Azyme, but instead on diffusive access to pA monomers, replication of the Azyme sequence by USR, and the spread of this sequence by polymer diffusion.

Functional selection. Of the four system parameters explored above, the monomer hopping rate, k_m, was selected for variation in a series of simulations in which the Azyme sequence "spontaneously" appears. This parameter was chosen because diffusion of pA monomers into the vicinity of an Azyme, as well as the diffusion of newly synthesized A monomers away from a Azyme, was expected to affect the ability for an Azyme to become established in a pre-existing population. In Figure 5, the Species Lifetime and Population Size of Azyme lineages are shown for simulations in which k_m was varied from 0.001 to 10. The polymer hopping rate, k_p, was fixed at $k_p = 0.01$, and all other parameters were the same as those used in the simulations presented in Figures 3 and 4. The selective advantage of the Azyme over nonfunctional sequences is clear in these simulations, with the lifetime of an Azyme sequence (shown in black) being 5 to 60 times as many cycles as the *average* lifetime of nonfunctional polymers (shown in green) in simulations with identical parameters, but without the appearance of an Azyme. The population size of the Azyme, for all k_m values explored, was approximately two times larger than the average population size of nonfunctional polymers in simulations where no Azyme appeared.

It is important to note that Species Lifetime and Population Size for simulations with nonfunctional polymers are weighted towards longer lifetimes and larger population sizes by sequences that appear early in the simulations. Early-time sequences are able to attain relatively large populations before the free monomer concentration begins to limit replication. Therefore, the average lifetimes and populations of early-time sequences are typically much greater in magnitude than those of nonfunctional sequences that emerge later, *e.g.* after DKS has been established. Thus, the observed enhanced species lifetime and population size of the late-appearing Azyme is even more significant than it first appears relative to the green curve in Figure 5. To illustrate this point, simulations were carried out in which a nonfunctional polymer was introduced at the same cycle time and lattice site as the Azyme (without the introduction of the Azyme). As shown by the blue curve in Figure 5A, the selective advantage of the Azyme sequence is, as expected, more dramatic when compared to the measured lifetimes and population sizes of the late-appearing nonfunctional sequence for all values of k_m.

The lifetime of an Azyme sequence appears to be less dependent on k_m than the non-functional sequences, being of similar lifetime from $k_m = 0.01$ to 10. One exception is the considerable increase in Azyme lifetime observed in simulations with $k_m = 0.001$. Under these conditions of very slow monomer diffusion, the lifetime of nonfunctional polymer sequences, even those that appear early in a simulation, are too short for any polymer lineages to become established and for a considerable population of any nonfunctional sequence to subsist. Thus, the results shown in Figure 5 also demonstrate that the Azyme is able to survive in an environment that cannot sustain nonfunctional polymer populations. In other words, the Azyme, by significantly enhancing its own survival, can become the first sustainable extant sequence in a highly dynamic and previously unsustainable environment.

System-level benefits of a functional sequence. The population of the Azyme is plotted in Figure 6A as a function of time (in units of cycles) for a simulation with $k_m = 10$. A comparison of this plot with those of simulations with only nonfunctional sequences reveals that the population growth of this sequence is faster and to a similar level of the most successful early-appearance polymers (*e.g.* Sequence ID 33 in Figure 1A). The effect of Azyme appearance on the overall system can be appreciated by comparing the total polymer population after the appearance of the Azyme to the steady-state polymer population that is maintained by nonfunctional sequences (Figure 6B). In these simulations, the Azyme becomes well established in the population, but its population growth represents only about 10% of the total increase in polymer population. Thus, most of the newly formed A monomers are used to generate new sequences and to replicate existing nonfunctional sequences. While this result might, at first, be considered deleterious for the Azyme and a waste of resources on "parasitic" nonfunctional sequences, it must be realized that the allocation of resources to other sequences is positive at a system level, as the capacity to search sequence space and to propagate other functional sequences is enhanced.

A comparison of $2D$ spatial maps of polymer and monomer densities for systems with and without the introduction of the Azyme sequence further illustrates the effects of a functional Azyme on a system of otherwise nonfunctional polymers (Figure 6C). The location of the initial appearance of the Azyme sequence is essentially devoid of polymers at $t = 5000$ cycles in the simulation containing only nonfunctional polymers. In contrast, introduction of the single Azyme sequence at $t = 2500$ cycles results in a substantial change in the local and global polymer distribution. A cluster grows around the site of Azyme introduction, which eventually merges with nearby clusters of nonfunctional sequences. Within this larger cluster the Azyme coexists with nonfunctional sequences that benefit from the temporal increase in local free A-monomer concentration. Dramatic differences in A-monomer and B-monomer densities are also observed across the simulation surface (Figure 6C). The A monomer is no longer a limiting factor in polymer production, and eventually increases to higher than pre-Azyme levels at every point on the simulation domain. In contrast, the B monomer becomes a more strongly limiting reagent to polymer production, with B-monomer levels dropping well below those of the pre-Azyme levels across the simulation surface as more resources are consumed by the larger polymer population.

Functional cooperation over time and space. Having demonstrated that a single functional sequence can become established within a system of nonfunctional polymers, we next investigated the effects of adding a second functional sequence that acts as a catalyst for the conversion of a proto-B monomer (pB) to B monomer. As shown in Figure 6A, in simulations where 60 pB monomers were initially present at each lattice site, the appearance of a single Bzyme sequence at $t = 4000$ cycles (1500 cycles after the appearance of the Azyme) quickly results in a burst of Bzyme population growth. As was observed for the Azyme, the growth in Bzyme population represents only a fraction of the total increase in the global polymer population size (Figure 6B). A plot of polymer density 1000 cycles after the appearance of the Bzyme shows the growth of a large cluster with the Bzyme population at its center. Other polymer clusters on the surface show substantial growth as a result of the appearance of the Bzyme.

An important feature of the system explored here is that the selective pressure for a functional polymer can be transient in time and space. Each simulation run shows slightly different dynamical evolution after the appearance of a functional polymer. Azyme and Bzyme population plots and $2D$ density plots for a second example are provided in the Supporting Information (Figure S6) for a simulation with $k_m = 1.0$ (as compared to $k_m = 10$ for the simulations shown in Figure 6). For this simulation where only the Azyme appears at 2500 cycles (the Bzyme does not appear), the Azyme sequence goes extinct within the next 6000 cycles. In contrast, when the Bzyme is nucleated near the center of Azyme activity, survival of both functional sequences is enhanced. We note that extinction of the Azyme, in the absence of Bzyme appearance, does not terminate system evolution. Figure S6B illustrates that the pool of polymers benefited from the transient activity of the Azyme. Furthermore, the Azyme had nearly exhausted its benefit to the system at the time of extinction, having had converted nearly all pA to A monomers. However, a stable cluster emerged where the Azyme nucleated (Figure S6C), leading to localized enhancement of polymer density and number of extant species. This result demonstrates that the Azyme (or any other functional sequence) is not required to live indefinitely in order to have a positive impact on a system undergoing continuous rounds of USR. Moreover, once the A-monomer is no longer a limiting reagent, it would be a distinct disadvantage for the Azyme population to remain high: for continued system-level evolution, it is more advantageous for the monomers in the Azyme sequences to be recycled into polymers with functions that are needed at later times.

Discussion

The physical environment and the molecules available on the prebiotic Earth would have placed tremendous constraints on *any* mechanism that led to the evolution of informational polymers with functional activity. Among these constraints would have been limited resources and finite polymer stability. With these particular constraints in mind, we used a kinetic Monte Carlo simulation to explore the emergence of functional polymers when only nonfunctional polymers existed that were all replicated with equal probability, regardless of sequence. The model utilizes a minimal set of adjustable parameters in order to explore the effects of those parameters considered most relevant to this putative early stage of prebiotic evolution. The results of our simulations have revealed the possible existence of regions in physiochemical parameter space that could have supported the constant exploration of sequence space, as well as the selection of functional sequences.

Our simulations demonstrate how variations in polymer hydrolysis and replication rates affect the ability for a pool of informational polymers to explore sequence space, even after an equilibrium population of polymers has been established. As expected, systems with faster polymer hydrolysis rates allow more rapid exploration of sequence space, due to more rapid turnover of resources. However, for a system to take advantage of functional sequences that appear spontaneously, the polymer replication rate must be sufficient to counter the deleterious effect of rapid polymer degradation, otherwise information propagation is not sustainable and no polymer lineages become established. Conversely, low polymer degradation rates can cause extant (and nonfunctional) polymers to unproductively retain monomer resources, severely limiting the rate of sequence space exploration. For the model parameters explored here, we have found that a region of compromise exists for a range of replication and hydrolysis rates that allows a nonzero rate of sequence exploration and a nonzero probability that new sequences become established in the system.

To demonstrate functional sequence selection we focused on the appearance and propagation of monomer synthetases. For this particular functionality, the appearance of the Azyme and Bzyme in the extant population at different points in space and time illustrates a rudimentary form of cooperativity between two functional lineages that may be spatially and/or temporally separated. We argue that such a scenario for functional sequence emergence and early sequence cooperation is more plausible for the earliest stages of prebiotic evolution than scenarios that require the first functional polymer to have been a much more complicated enzyme (*i.e.* a processive polymer replicase [54]), or for the diverse members of a set of enzymes to emerge *de novo* at the same point in space and time [68]. As the system evolved, there is no reason not to expect that continued system dynamics would eventually permit more complicated catalytic functionalities to be selected and optimized, perhaps even culminating in the eventual appearance of polymerases [69] and ligases [70].

Many origin of life researchers consider polymer compartmentalization, such as nucleic acid encapsulation in lipid vesicles [71], to be a prerequisite for evolution. It is certainly true that there must exist a means for the co-localization of functional polymers with the "fruits of their labor". However, limited diffusion (or incomplete mixing) has been shown to provide a possible alternative to encapsulation [9,32,36], at least in the earliest stages of prebiotic evolution. As shown here, system-level metrics, such as Average Species Size and Exploration of Sequence Space, depend on the rates of molecule movement between lattice sites, illustrating that limited movement, representing a realm between stringent compartmentalization and homogenous mixing, could have been beneficial in the early stages of informational polymer growth and evolution. Furthermore, even without employing explicit compartmentalization, nearly all diffusion-limited regimes explored here support stable and diverse populations of extant sequences, and some promote the dynamic emergence of spatial aggregates, which appear even in the absence of any functional activity. In particular, our simulations illustrate how nonfunctional informational polymers can play a positive role by contributing to a local recycling dynamic that sustains extant polymer populations. As one consequence, in all diffusive regimes explored, the unit of selection (or survival) is not the individual polymer, but local populations of polymers dynamically coupled through resource recycling. These populations act both cooperatively as competitive aggregates and individually through single polymer replication/degradation/diffusion events, and through functional selection of active sequences. Thus, the results presented here reinforce previous assertions that physical compartmentalization is not necessary for prebiotic evolution [9, 36].

The effects of parasites are not absent in our models. For example, in simulations where *A*zyme and *B*zyme sequences emerge, the majority of monomer resources created by these functional sequences become incorporated into nonfunctional sequences. The observed dynamics are similar to that of the nonfunctional parasites in the model of Könnyű and Czárán [37]. However, in contrast to the dynamics observed in their metabolic autonomous replicator model, where parasites are tolerated by functional sequences but play no active role, parasites in the prebiotic scenario presented here are not completely deleterious. When a synthetase first emerges nonfunctional sequences may be beneficial by providing a localized enhancement of resources in an existing polymer cluster that locally retains the new monomer resources (within recyclable polymers) as the synthetase sequence gradually increases in number. At the very least, nonfunctional sequences that take up monomers generated by the synthetase can later provide raw materials for the continued search of sequence space, thereby increasing the survivability and evolvability of the system as a whole.

In conclusion, we have shown that system-level phenomena can emerge from a pool of monomers and replicating polymers that are governed by a small number of meaningful chemical and physical parameters. Moreover, we have shown that in a system of polymers where there is no intrinsic sequence-specific replicative advantage, evolution can still take place. At the rudimentary level – before the appearance of functional sequences – environmental cycling, limited diffusivity, and resource limitation leads to the spontaneous self-organization of polymers into spatial aggregates. Even in the idealized case of universal sequence replication, a dynamic fitness landscape spontaneously appears. When functional sequences eventually appear, they can become established amid the nonfunctional polymers, and enhance the ability for other sequences to evolve across space and time. Taken together, these results allow qualitative predictions about the chemistries and environments that would have facilitated prebiotic sequence evolution. Specifically, our model suggests that the optimal conditions for the earliest stage of abiotic chemical evolution, prior to the onset of functional evolution, would *not* have been those that promote stringent maintenance of sequence information *per se*, which is considered optimum for most autonomous replicator models [54], including that of Eigen [6, 8]. On the contrary, the optimum conditions for early informational polymer evolution would have allowed the spontaneous appearance of completely new sequences, and for the existence of new sequences to be just long enough for functional sequences to gain a local selective advantage during subsequent cycles of replication. Future investigations of the earliest replicating polymers of life should therefore place more emphasis on polymer chemistries and environments that allow for rapid turnover of resources and time-varying diffusivities for monomers and polymers. Finally, in light of the possibility that compartmentalization appeared after the onset of functional polymer evolution, the results presented here support a model for the early stages of biopolymer evolution that are dramatically different from that governed by a strictly Darwinian process. That is, these early stages could have been defined by the collective evolution of a system-wide cooperation of polymer aggregates. The same general characteristics have been proposed by Woese for the earliest biological systems, but for reasons based on bioinformatics analyses of extant organisms [72, 73].

Acknowledgments

We thank Profs. Frank A. L. Anet and Loren Dean Williams for helpful discussions. This work was jointly supported by NSF and the NASA Astrobiology Program, under the NSF Center for Chemical Evolution, CHE-1004570. SIW additionally acknowledges support from the NASA Astrobiology Institute through the NASA Postdoctoral Fellowship Program.

References

1. Szathmáry E, Smith JM (1997) The major transitions in evolution. Oxford: Oxford University Press.

2. Kacian D, Mills D, Kramer F, Spiegelman S (1972) A replicating RNA molecule suitable for a detailed analysis of extracellular evolution and replication. Proc Natl Acad Sci USA 69: 3038–3042.

3. Ellington A, Szostak J (1990) In vitro selection of RNA molecules that bind specific ligands. Nature 346: 818–822.

4. Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249: 505–510.

5. Wochner A, Attwater J, Coulson A, Holliger P (2011) Ribozyme-catalyzed transcription of an active ribozyme. Science 332: 209–212.

6. Eigen M (1971) Self–organization of matter and evolution of biological macromolecules. Naturwissenschaften 58: 465–523.

7. Eigen M, Schuster P (1977) A principle of natural self-organization. Naturwissenschaften 64: 541–565.

8. Eigen M, McCaskill J, Schuster P (1988) Molecular quasi-species. J Phys Chem 92: 6881–6891.

9. Szabo P, Scheuring I, Czárán T, Szathmáry E (2002) In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. Nature 420: 340–343.

10. Joyce GF (2002) Molecular evolution: Booting up life. Nature 420: 278–279.

11. Joyce GF (2002) The antiquity of RNA-based evolution. Nature 418: 214–221.

12. Wu M, Higgs P (2009) Origin of self-replicating biopolymers: Autocatalytic feedback can jump-start the RNA world. J Mol Evol 69: 541–554.

13. Gesteland R, Cech T, Atkins JF, editors (2006) The RNA World. Cold Spring harbor, NY: Cold Spring Harbor Lab. Press, 3rd edition.

14. Nielsen PE (1993) Peptide nucleic acid (PNA): A model structure for the primordial genetic material? Orig Life Evol B 23: 323–327.

15. Eschenmoser A (2007) The search for the chemistry of life's origin. Tetrahedron 63: 12821–12844.

16. Engelhart AE, Hud NV (2010) Primitive genetic polymers. Cold Spring Harb Perspect Biol 2.

17. Lee DH, Granja J, Martinez J, Severin K, Ghadiri M (1996) A self-replicating peptide. Nature 382: 525–528.

18. Cairns-Smith AG (1982) Genetic takeover and the mineral origins of life. Cambridge: Cambridge University Press.

19. Robertson MP, Miller SL (1995) An efficient prebiotic synthesis of cytosine and uracil. Nature 375: 772–774.

20. Commeyras A, Collet H, Boiteau L, Taillades J, Vandenabeele-Trambouze O, et al. (2002) Prebiotic synthesis of sequential peptides on the Hadean beach by a molecular engine working with nitrogen oxides as energy sources. Polym Int 665: 661–665.

21. Hud NV, Anet FL (2000) Intercalation-mediated synthesis and replication: A new approach to the origin of life. J Theor Biol 205: 543–562.

22. Corliss JB, Baross JA, Hoffman SE (1981) A hypothesis concerning the relationship between submarine hot springs and the origin of life on earth. Oceanol Acta 4: 59–69.

23. Huber C, Wächtershäuser G (1998) Peptides by activation of amino acids with CO on (Ni,Fe)S surfaces: Implications for the origin of life. Science 281: 670–672.

24. Wächtershäuser G (1992) Groundworks for an evolutionary biochemistry: the iron-sulphur world. Prog Biophys Mol Biol 58: 85–201.

25. Hazen RM, Sverjensky DA (2010) Mineral surfaces, geochemical complexities, and the origins of life. Cold Spring Harb Perspect Biol 2.

26. Stribling R, Miller SL (1991) Template-directed synthesis of oligonucleotides under eutectic conditions. J Mol Evol 32: 289–295.

27. Monnard PA, Kanavarioti A, Deamer DW (2003) Eutectic phase polymerization of activated ribonucleotide mixtures yields quasi-equimolar incorporation of purine and pyrimidine nucleobases. J Am Chem Soc 125: 13734–13740.

28. Monnard PA, Ziock H (2008) Eutectic phase in water-ice: A self-assembled environment conducive to metal-catalyzed non-enzymatic rna polymerization. Chem Biodivers 5: 1521–1539.

29. Pedersen K (2000) Exploration of deep intraterrestrial microbial life: current perspectives. FEMS Microbiol Lett 185: 9–16.

30. Dobson CM, Ellison GB, Tuck AF, Vaida V (2000) Atmospheric aerosols as prebiotic chemical reactors. Proc Natl Acad Sci USA 97: 11864–11868.

31. Czárán T, Szathmáry E (2000) Coexistence of replicators in prebiotic evolution. In: U Dieckmann RL, Metz JAJ, editors, The Geometry of Spatial Interactions: Simplifying Spatial Complexity, Cambridge University Press.

32. Hogeweg P, Takeuchi N (2003) Multilevel selection in models of prebiotic evolution: Compartments and spatial self-organization. Orig Life Evol B 33: 375-403.

33. Ma W, Yu C, Zhang W (2007) Monte carlo simulation of early molecular evolution in the RNA world. Biosystems 90: 28–39.

34. Ma W, Yu C, Zhang W, Hu J (2007) Nucleotide synthetase ribozymes may have emerged first in the RNA world. RNA 13: 2012–2019.

35. Könnyű B, Czárán T, Szathmáry E (2008) Prebiotic replicase evolution in a surface-bound metabolic system: parasites as a source of adaptive evolution. BMC Evol Biol 8: 267.

36. Takeuchi N, Hogeweg P (2009) Multilevel selection in models of prebiotic evolution II: A direct comparison of compartmentalization and spatial self- organization. PLoS Comput Biol 5: e1000542.

37. Könnyű B, Czárán T (2011) The evolution of enzyme specificity in the metabolic replicator model of prebiotic evolution. PLoS ONE 6: e20931.

38. Zhan ZYJ, Lynn DG (1997) Chemical amplification through template-directed synthesis. J Am Chem Soc 119: 12420–12421.

39. Leitzel J, Lynn DG (2001) Template-directed ligation: From DNA towards different versatile templates. Chem Rec 1: 53–62.

40. Bean HD, Anet FAL, Gould IR, Hud NV (2006) Glyoxylate as a backbone linkage for a prebiotic ancestor of RNA. Orig Life Evol B 36: 39–63.

41. Hud NV, Jain SS, Li X, Lynn DG (2007) Addressing the problems of base pairing and strand cyclization in template-directed synthesis - a case for the utility and necessity of “molecular midwives” and reversible backbone linkages for the origin of proto-RNA. Chem Biodivers 4: 768–783.

42. Ura Y, Beierle JM, Leman LJ, Orgel LE, Ghadiri MR (2009) Self-assembling sequence-adaptive peptide nucleic acids. Science 325: 73–77.

43. Li X, Hernandez AF, Grover MA, Hud NV, Lynn DG (2011) Step-growth control in template-directed polymerization. Heterocycles 82: 1477–1488.

44. Lahav N, White D, Chang S (1978) Peptide formation in prebiotic era - thermal condensation of glycine in fluctuating clay environments. Science 201: 67–69.

45. Apel CL, Deamer DW (2005) The formation of glycerol monodecanoate by a dehydration/condensation reaction: Increasing the chemical complexity of amphiphiles on the early earth. Orig Life Evol B 35: 323–332.

46. Hazen RM (2009) The emergence of patterning in life’s origin and evolution. Int J Dev Biol 53: 683–692.

47. Fishkis M (2011) Emergence of self-reproduction in cooperative chemical evolution of prebiological molecules. Orig Life Evol B 41: 261–275.

48. King GAM (1982) Recycling, reproduction, and life’s origins. Biosystems 15: 89–97.

49. King GAM (1986) Was there a prebiotic soup? J Theor Biol 123: 493–498.

50. Chacón P, Nuño JC (1995) Spatial dynamics of a model for prebiotic evolution. Physica D 81: 398–410.

51. Deck C, Jauker M, Richert C (2011) Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. Nature Chem : 1–6.

52. Luther A, Brandsch R, von Kiedrowski G (1998) Surface-promoted replication and exponential amplification of DNA analogues. Nature 396: 245–248.

53. Boerlijst M, Hogeweg P (1991) Spiral Wave Structure in pre-biotic evolution: Hypercycles stable against parasites. Physica D 48: 17–28.

54. Eors, Szathmáry (1989) The integration of the earliest genetic information. Trends Ecol Evol 4: 200–204.

55. Lehman N, Arenas CD, White WA, Schmidt FJ (2011) Complexity through recombination: From chemistry to biology. Entropy 13: 17–37.

56. Hazen RM, Griffin PL, Carothers JM, Szostak JW (2007) Functional information and the emergence of biocomplexity. Proc Natl Acad Sci USA 104: 8574-8581.

57. Horowitz ED, Engelhart AE, Chen MC, Quarles KA, Smith MW, et al. (2010) Intercalation as a means to suppress cyclization and promote polymerization of base-pairing oligonucleotides in a prebiotic world. Proc Natl Acad Sci USA 107: 5288–5293.

58. Jain SS, Anet FAL, Stahle CJ, Hud NV (2004) Enzymatic behavior by intercalating molecules in a template-directed ligation reaction. Angew Chem Int Edit 43: 2004–2008.

59. Alfonsi A, Cancès E, Turinici G, Di Ventura B, Huisinga W (2004) Exact simulation of hybrid stochastic and deterministic models for biochemical systems. Rapport de recherche RR-5435, INRIA.

60. Chatterjee A, Vlachos DG (2007) An overview of spatial microscopic and accelerated kinetic Monte Carlo methods. J Comput-Aided Mater 14: 253–308.

61. Turk RM, Chumachenko NV, Yarus M (2010) Multiple translational products from a five-nucleotide ribozyme. Proc Natl Acad Sci USA 107: 4585–4589.

62. T D, Gillespie (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J Comput Phys 22: 403–434.

63. Manapat ML, Chen IA, Nowak MA (2010) The basic reproductive ratio of life. J Theor Biol 263: 317–327.

64. Bernstein D (2005) Simulating mesoscopic reaction-diffusion systems using the Gillespie algorithm. Phys Rev E 71: 41103.

65. Pross A (2003) The driving force for life's emergence: Kinetic and thermodynamic considerations. J Theor Biol 220: 393–406.

66. Pross A (2005) On the emergence of biological complexity: Life as a kinetic state of matter. Orig Life Evol Biosph 35: 151–166.

67. Shannon C (1948) A mathematical theory of communication. Bell Syst Tech J 27: 623–656.

68. Kauffman S (1993) The Origins of Order: Self-organization and Selection in Evolution. Oxford: Oxford University Press.

69. Wu M, Higgs PG (2011) Comparison of the roles of nucleotide synthesis, polymerization, and recombination in the origin of autocatalytic sets of RNAs. Astrobiology 11: 895–906.

70. Ma W, Yu C, Zhang W, Hu J (2010) A simple template-directed ligase ribozyme as the RNA replicase emerging first in the RNA world. Astrobiology 10: 437–447.

71. Szathmáry E, Demeter L (1987) Group selection of early replicators and the origin of life. J Theor Biol 128: 463–486.

72. Woese CR (2002) On the evolution of cells. Proc Natl Acad Sci USA 99: 8742–8747.

73. Vestigian K, Woese C, Goldenfeld N (2006) Collective evolution and the genetic code. Proc Natl Acad Sci USA 103: 10696–10701.

74. Press W, Teukolsky S, Vetterling W, Flannery B (1992) Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, 2nd edition.

75. Filotas E, Grant M, Parrott L, Arne P (2010) Positive interactions and the emergence of community structure in metacommunities. J Theor Biol 266: 419–429.

76. Chao A, Chazdon RL, Colwell RK, Tsung-Jen S (2005) A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecol Lett 8: 148–159.

A Supporting Information: Methods S1

A.1 Computational Methods

Simulations of the spatiotemporal dynamics of diffusion-limited informational polymers with environmentally driven assembly (both spontaneous and template-directed) and degradation were implemented via a hybrid kinetic Monte Carlo algorithm [59,60], as outlined here.

A.1.1 Model Definition

Our model describing the dynamics of environmentally-driven recycling of informational polymers includes three kinetic and two diffusive processes:

1. Sequence-independent spontaneous assembly of polymers, with a rate constant k_s
2. Template-directed polymer replication via Universal Sequence Replication (USR), with a rate constant k_r
3. Sequence-independent polymer hydrolysis, with a rate constant k_h
4. Monomer diffusion, with hopping rate k_m (related to monomer diffusivity $\mathcal{D}_m$ by eq. 23)
5. Polymer diffusion, with hopping rate k_p (related to polymer diffusivity $\mathcal{D}_p$ by eq. 23).

The kinetic rate constants k_s, k_r, k_h, and diffusive hopping rate constants k_m and k_p are the tunable parameters in our model. Throughout we use italic and lowercase letters to denote dimensionless simulation values and upper-case letters to indicate the physical values of parameters. Since our aim is to study the influence of kinetic and physical parameters on system dynamics, additional model parameters, including total system mass and cycling rate, are held at a fixed values in this work. This simplified model permits us to explore a wide variety of chemistry-environmental couplings (*i.e.* a wide range of kinetic and diffusive parameter space) with the tractable set of just five model parameters.

The corresponding dimensionless mass-action kinetic equations for our model system are:

$$\frac{\partial A}{\partial t} = \mathcal{D}_m \nabla^2 A - R_{L_{1/2}} k_s AB - R_{L_{1/2}} k_r AB \sum_i X_i + R_{L_{1/2}} k_h \sum_i X_i \quad (6)$$

$$\frac{\partial B}{\partial t} = \mathcal{D}_m \nabla^2 B - R_{L_{1/2}} k_s AB - R_{L_{1/2}} k_r AB \sum_i X_i + R_{L_{1/2}} k_h \sum_i X_i \quad (7)$$

$$\frac{\partial X_i}{\partial t} = \mathcal{D}_p \nabla^2 X_i + k_s AB + k_r ABX_i - k_h X_i \quad (8)$$

where $R_{L_{1/2}}$ is half of the polymer length R_L (the motivation for including this factor is outlined below), and the sum over i runs over all extant sequences (*i.e.* all unique species in the population at time t). The variables A, B, and X_i correspond to the dimensionless concentrations of A monomer, B monomer, and polymer species with $ID = i$ respectively. This choice of kinetics is intended to approximate spontaneous assembly as a nucleation event, where the potential barrier for nucleation of a new sequence is high but once surmounted the sequence is easily assembled (see Section 2.2.4 in main text for more discussion). Therefore, the rate of spontaneous assembly is governed by the dimensionless second-order rate constant k_s and the local abundances $A(x, y)$ and $B(x, y)$. Likewise, replication is modeled as a third-order process (describing nucleation on a template), governed by a dimensionless sequence-independent third-order rate constant k_r, the local monomer concentrations $A(x, y)$ and $B(x, y)$, and the local abundance of polymer species i, X_i. Hydrolysis is governed by the first-order rate constant k_h and species abundance (see Section 2.2.4 in main text for more discussion). Length dependence is introduced to the kinetic equations through the factor $R_{L_{1/2}}$. The system has a conserved total mass, *i.e* we study a closed-mass system. Defining M as the total mass (the total number of monomeric units in monomer and as polymeric residues), we require $M = R_L \sum_i X_i + A + B$.

The number of possible polymer sequences N increases exponentially with the polymer length R_L; we therefore simplify our model by limiting the space of possible sequences through constraining the number of monomer species, the length of polymers, and the possible sequence content of polymers. We implement our model with two monomer species, labeled A and B. This choice is motivated by selecting the minimal system that will permit diversity of polymeric sequences. We consider only polymers with length $R_L = 20$, and each distinct polymeric sequence contains both 10 A-monomers and 10 B-monomers. This simplifies polymer identification: each unique species need only be identified by a single ID number, i, where the ID number contains enough information to fully identify the species. Our approximation to the full sequence space greatly simplifies the model while maintaining the essential dynamical features we wish to capture (see discussion in Section in 2.2.2 and 2.2.3 of the main text). It also justifies our implementation of the factor $R_{L_{1/2}}$ included in eqs. 6 - 8, because both monomer species A and B contribute $R_{L_{1/2}}$ residues to a given polymer. Additionally, the fixed polymer length of $R_L = 20$ was chosen to balance our requirements of having sufficient diversity in possible polymer sequences while keeping the dynamics numerically tractable. The number of possible polymers for $R_L = 20$ with a $50 : 50$, $A : B$ ratio is calculated by the binomial coefficient, $20!/(10!10!) = 184,756$. This set is large enough such that the simulations presented here only sample a small fraction (less than 5%) of the total number of possible sequences available in the potential sequence pool. The fixed value $R_L = 20$ can be taken to approximate a reaction-diffusion system supporting polymers with a mean-length of $\bar{L} = 20$ residues. These simplifications greatly alleviate the numerical requirements of our simulations, allowing us to run larger statistical samplings of our *in silico* experimental runs, while still permitting us to explore the most salient features of system dynamics. We have also explored systems supporting polymerization of shorter sequences, with $R_L < 20$, and observed that the dynamics are qualitatively similar, where the ratio $\frac{N_m}{N_p}$ scales with polymer length for a given set of kinetic parameters (N_m and N_p are the total number of monomers and polymers, respectively).

A.1.2 Model Implementation

The reaction surface is modeled on a 64×64 square lattice. The lattice size of 4096 sites was chosen to be small enough for numerical tractability, yet large enough to allow spatial correlations to be observed, for the ranges of kinetic and diffusive parameter under study. The initial conditions are homogeneous, with all mass in monomer: each of the 4096 lattice sites is initialized with 60 A and 60 B monomers. For functional runs each site is additionally initialized with 60 pA monomers. No polymers are present at the start of a given simulation run (with the exception functional sequence runs which start from an already established pool of random sequence polymers at quasi steady-state, see below). Periodic boundaries were imposed to avoid edge-effects. However, additional simulations performed with closed boundaries revealed similar dynamics to those presented here. For example, in the case of simulation parameters leading to the emergence of localized clusters of high polymer density, we observe polymers preferentially aggregating at the corners of the square grid (not shown).

Each of the 4096 lattice sites on the two-dimensional grid represents a locally homogenous reaction domain, characterized by a locally interacting community of monomers and polymers of different sequences. All kinetic events occur locally (on-site). Interaction between lattice sites occurs through diffusive contact. Over the course of system evolution, large separations in polymer and monomer populations are observed, with a typical simulation ratios (at steady-state) of $\sim \frac{N_m}{N_p} = 300$, where N_m and N_p are the total numbers of monomers and polymers, respectively. To increase numerical efficiency, reaction and diffusion events are therefore partitioned such that relatively rare polymer reaction and diffusion events are treated stochastically and much more frequent monomer diffusion events are treated deterministically. This partitioning is implemented with a spatially explicit hybrid kinetic Monte Carlo algorithm [59, 60] (see below). We have verified that the fully stochastic spatially-explicit kinetic Monte Carlo simulation [64] quantitatively reproduces the same dynamical phenomena (not shown).

The kinetic equations outlined in eqs. 6 – 8 do not explicitly account for environmental cycling. However, our numerical implementation via kinetic Monte Carlo simulations must explicitly take into account environmental cycling in order to accurately capture the desired dynamics. Our Monte Carlo implementation is therefore partitioned into two phases: a dehydrated phase, where all sites are diffusively isolated (*i.e.* diffusion is turned off); and a hydrated phase, where sites interact diffusively through polymer hopping events and monomer diffusion. The three kinetic processes are partitioned between these two phases such that polymer assembly and replication occur in the dehydrated phase, and polymer degradation and enzymatic catalysis (when functional sequences are extant) occur in the hydrated phase.

The Dehydrated Phase Each lattice site x on the two-dimensional lattice is diffusively isolated and treated individually with the standard Gillespie algorithm [62]. The dehydrated phase supports two kinetic processes:

- Spontaneous assembly, with propensity:

$$a_s(x) \quad = \quad k_s A(x) B(x) \tag{9}$$

- Sequence-independent replication via template-directed assembly, with propensity:

$$a_{r_i}(x) \quad = \quad k_r A(x) B(x) N_i(x) \tag{10}$$

Here $A(x)$, $B(x)$, and $N_i(x)$ are the *number* of A monomers, B monomers, and polymer species i at site x (which is not strictly the same as the dimensionless concentrations cited for eqs. 6 – 8 above). The total propensity for polymer formation is

$$a_{tot}(x) \quad = \quad a_s(x) + \sum_i a_{r_i}(x) \ , \tag{11}$$

where the sum over i runs over all species indigenous to the site x. Events are drawn at random. Since the dynamics are environmentally driven, a polymer may copy itself *at most* once per hydration/dehydration cycle. Therefore, after each replication event the total number of polymers available for replication is decreased by one unit. As A and B monomers are incorporated into polymers over the course of the dehydrated phase, the propensities for polymer formation through replication and spontaneous assembly decrease.

For all simulations presented in this work, the rate constant for spontaneous assembly is set to the dimensionless value $k_s = 10^{-7}$. This value corresponds to a maximal global nucleation rate of 1.47 new sequences per environmental cycle when no polymers are present (calculated by summing $a_s = k_s \times AB$ over all sites for the case where all mass is in monomer, *i.e.* $A = B = 60$ monomers per site for all 4096 lattice sites).

The Hydrated Phase Diffusion and hydrolysis occur when the system is hydrated. In the hydrated-phase the dynamics are modeled with a spatially-explicit hybrid kinetic Monte Carlo algorithm [59, 60], where polymer diffusive hopping is treated as a stochastic event as in other spatially explicit treatments which incorporate diffusion in the Gillespie algorithm (see *e.g.* [64]). For the system presented here, a natural partition between rare and common events occurs due to the large separation in the population densities of monomer and polymer as discussed above. Our simulations are therefore hybridized such that monomer site hopping is coarse-grained by introducing the partial differential equation describing mass-action monomer diffusion:

$$\frac{\partial M(x)}{\partial t} \quad = \quad \mathcal{D}_m \nabla^2 M(x) \tag{12}$$

where $\mathcal{D}_m$ is the macroscopic monomer dimensionless diffusion rate (see eq. 19 below for relation to monomer hopping rate cited in the text), $M(x) = A(x)$ or $B(x)$ is the concentration of monomers per unit area, and ∇^2 is the two-dimensional Laplace operator. Eq. 12 is evolved forward in time using a standard finite-difference staggered leapfrog algorithm [74], with lattice spacing $dx = 0.2$ and time-step $dt = 0.01$. Implementation of this algorithm is subject to the stability criteria imposed by the Courant Condition yielding numeric stability for $\mathcal{D}_m < 1$, constraining our simulations to values $k_m < 100$ (see eq. 19), for the values of dx and dt implemented in the simulations.

Against the background of monomer diffusion, the rare polymer events of diffusion and degradation occur. These rare processes are:

- Polymer hydrolysis (degradation) with propensity

$$a_{h_i}(x) \quad = \quad k_h N_i(x) \tag{13}$$

- Polymer diffusion (hopping between nearest-neighbor lattice sites) with propensity

$$a_{p_i}(x) \quad = \quad k_p N_i(x) \tag{14}$$

The time-step between rare-events is calculated by determining the value of τ such that

$$\int_t^{t+\tau} a_{rare}(t')dt' = \ln\left(\frac{1}{\xi}\right) \tag{15}$$

where a_{rare} is the total propensity for rare reactions,

$$a_{rare} = \sum_x \sum_i a_{p_i}(x) + a_{h_i}(x) \ , \tag{16}$$

with events selected globally, and ξ is a random number drawn from the set $(0, 1]$ [59]. To calculate the integral in eq. 15, the numbers of all rare and common molecules must be followed in time. The deterministic PDE of eq. 12 is used to track the evolution of the monomers species A and B. The PDEs are integrated forward in time using the staggered leapfrog algorithm until eq. 15 is satisfied. One stochastic event is then chosen to occur at random from the global pool of possible rare events on the lattice by a standard Gillespie algorithm [62], which is calculated only for rare events.

Functional Runs In simulations including a functional Azyme, which catalyzes formation of A monomers from pA monomers, two additional processes are added to the hydrated phase: diffusion of pA monomers, and catalytic conversion of $pA \rightarrow A$. Diffusion of pA monomers is treated the same as A– and B– monomer diffusion using the mass-action monomer diffusion eq. 12. Catalysis is added to the rare events of the hydrated phase, with the probability of catalysis calculated from the reaction propensity:

$$a_{pA}(x) \quad = \quad k_c pA(x) N_j(x) \tag{17}$$

where $pA(x)$ is the dimensionless pA concentration (*i.e.* the number of pA monomers at site x), k_c is the microscopic catalysis rate for conversion of $pA \rightarrow A$ in the presence of the Azyme, and j is the ID number of the Azyme. The total propensity for rare events with an extant functional Azyme is therefore modified such that $a_{rare} = \sum_x (a_A(x) + \sum_i (a_{p_i}(x) + a_{h_i}(x))$.

As discussed in the main text, to illustrate the impact of the emergence of a functional sequence, data was saved at $t = 2500$ cycles for all details of a simulation, before continuing to run the simulation with no functional sequences present. This data provided the initial starting distribution of monomers and polymer species for the functional runs. The choice of starting at $t = 2500$ cycles is somewhat arbitrary, but was chosen to be sufficiently late in the system evolution that a quasi-steady state had been established (*i.e.* the ratio of monomer to polymer was relatively constant). To this initial condition, 60 pA monomers were added to each site (to model a previously untapped resource in the environment). A single polymer sequence representing the Azyme was randomly inserted on the lattice as a spontaneous assembly event (with the choice of site weighted by the propensities in eq. 9 for each run). Catalytic efficiency, k_c, was set to 100. Results were averaged over twenty-five runs (each with a randomly chosen insertion point for the inoculated sequence). The simulations were permitted to run until the inoculated sequence died out, or until the sequence had survived for 5000 cycles. The Bzyme was later inserted at $t = 4000$ cycles, catalyzing $pB \rightarrow B$.

A.2 Dimensionalization of Model Parameters

Throughout this work, we cite dimensionless microscopic kinetic and diffusive rates used in the kinetic Monte Carlo simulations. To recover dimensioned values we define the dimensionless time and space variables

$$t = \frac{T}{\gamma} \ , \qquad x = \frac{X}{\lambda} \ , \tag{18}$$

where γ is a typical time scale, and $\lambda = \sqrt{D\gamma}$ is a typical spatial scale, and D is a dimensioned diffusion rate which we scale out of the equations such that:

$$\mathcal{D}_m = \frac{D_m}{D} \ , \qquad \mathcal{D}_p = \frac{D_p}{D} \tag{19}$$

The system is subject to the physical constraint $\frac{D_p}{D_m} \geq 1$. With this parameterization, the dimensionless kinetic rate constants implemented in our simulations (indicated by lower-case letters) are related to their physical counterparts (denoted by upper-case letters) as:

$$k_h = \gamma K_h \quad , \quad k_s = \frac{\gamma K_s}{\lambda^2} \quad , \quad k_A = \frac{\gamma K_A}{\lambda^2} \quad , \quad k_r = \frac{\gamma K_r}{\lambda^4} \tag{20}$$

for hydrolysis (first-order process), spontaneous assembly and enzymatic catalysis (second-order processes), and templated-directed replication (third-order process), respectively. Dimensionless concentrations may be made dimensional by the transformation:

$$A = [A]\lambda^2 \ , \qquad B = [B]\lambda^2 \ , \qquad N_i = [X_i]\lambda^2 \tag{21}$$

where $[A]$ is the concentration of A monomers, $[B]$ is the concentration of B monomers, and $[X_i]$ is the concentration of polymer species with ID number i. The notation $[\ldots]$ is used to indicate concentration in number of molecules per unit area such that the variables A, B, and N_i correspond to the number of molecules within a spatial region of size λ^2. As shown below, it is convenient to take the length scale λ to be the physical size of a lattice site.

To model diffusion events on the lattice, a relation between the mass-action diffusivities defined above, and the microscopic hopping rates must be defined. The microscopic diffusive hopping rate for polymers is defined as:

$$k_p = \frac{4\mathcal{D}_p}{dx^2} \tag{22}$$

where we have used the mean-square displacement for particles subject to Brownian diffusion $\langle r^2 \rangle = 2d\mathcal{D}_p\tau$, with dimensionality $d = 2$, and a mean displacement $\langle r^2 \rangle = dx^2$, noting that $\tau = k_p^{-1}$ is the hopping time between sites. It is the dimensionless hopping rate k_p which we use as a simulation parameter and therefore cite throughout this work. The dimensional diffusivity D_p can be recovered using eqs. 19 and 22.

Since monomer diffusion is treated deterministically via a mass-action diffusion equation, the monomer diffusion parameter used in our simulations is $\mathcal{D}_m$. However, we additionally define a dimensionless monomer hopping rate:

$$k_m = \frac{4\mathcal{D}_m}{dx^2} \tag{23}$$

which is calculated in an analogous manner to k_p. Throughout this work we cite the dimensionless hopping rate k_m, rather than the simulation parameter $\mathcal{D}_m$, to provide an more direct comparison between monomer and polymer diffusion rates and timescales.

For numerical simplicity, we normalize the physical size of a lattice site L, to $L^2 = 1\ l^2$, where l is a unit of length (*e.g.* $l = \mu$m, mm, or cm), such that the number of molecules on a lattice site is equal to its dimensionless concentration (*i.e.* $N_i = X_i$ where X_i is the dimensionless concentration of species i cited in eqs. $6-8$). Consequently, the non-dimensional size of a lattice site is $dx = 0.2$ for our simulations, and using eq. 18 this yields a constraint condition

$$\sqrt{D\gamma} = 5\ l, \tag{24}$$

(where we have used $\lambda = \sqrt{D\gamma} = \frac{L}{dx}$), in our model parameterization. Scaling parameters D and γ must be chosen to satisfy eq. 24, to be consistent with our numerical implementation. Because it is easiest to define a timescale γ relative to the typical physical length of an environmental cycle, this usually translates to a constraint on the scaling parameter D, which is otherwise an arbitrary parameter. This choice therefore does not affect interpretation of our results, and enables straightforward recovery of dimensionalized values for simulation parameters by defining typical length and timescales λ and γ only.

A.3 Data Analysis

In the absence of functionality, the system comes to a quasi steady-state of dynamic kinetic equilibrium after several hundred cycles. The exact timescale depends on the particular diffusive and kinetic rates of a given system. Equilibrium is defined by a quasi steady-state ratio of monomer to polymer. That is, $\frac{d(N_m/N_p)}{dt} \simeq 0$, where N_m is the total number of monomers and N_p is the total number of polymers, with stochastic fluctuations about the equilibrium value. The quasi steady-state distribution is dynamic, with the population of extant polymers continuously changing with time. Nonetheless, certain global measurements, including the total number of polymers, the global and averaged local diversity (defined below), the average number of extant species, and the average extant species population size and lifetime, reach values during the quasi steady-state that are characteristic of a given set of system parameters (*i.e.* these values are deterministic, with stochastic fluctuations about a characteristic value). To make comparisons between different systems, with different sets of kinetic and diffusive parameters, we time-averaged these characteristic values during the quasi-steady state system evolution. Data is time-averaged over 2500 cycles and then ensemble averaged over a small statistical sampling of runs (for work presented here these sample sizes are 5 or 10 runs). Small statistical samples are sufficient given the small spread in simulation values and the length of simulations with time-sampling over 2500 cycles. The quasi-steady state distribution values are calculated starting at $t = 2500$ cycles, to ensure that all systems have achieved a steady-state distribution of monomer to polymer for all parameter ranges explored (typically steady-state is achieved at $t = 500 - 1000$ cycles depending on simulation parameters). Error bars correspond to sample standard deviation on the mean time-averaged values.

For runs illustrating the emergence of a functional sequence, where a random sequence is inoculated at a randomized location on the lattice (weighted by the propensities for spontaneous assembly at each site), lifetimes are averaged over survival times for the inoculated sequence lineage taken over all twenty-five experimental runs. Population size averages, as well as the average number of extant species and exploration rate, are averaged over the sequence lifetime (*i.e.* for a polymer that lives 5 cycles this corresponds to an average over 5 cycles) and therefore yield much higher variances in the data set than for the nonfunctional simulations which are averaged over entire populations of thousands of sequences, over thousands of environmental cycles.

A.3.1 Local Diversity

Each lattice site is characterized by its sequence population. To calculate the local (on-site) diversity of polymer populations we use the Shannon equation for information [67], used in studies of species diversity (see *e.g.* [75]). Local diversity is calculated for each site individually as,

$$S_L(x,t) = -\sum_i \frac{N_i(x,t)}{N_{tot}(x,t)} \ln\left(\frac{N_i(x,t)}{N_{tot}(x,t)}\right) \tag{25}$$

where i sums over all unique polymer species on the local site x, $N_i(x,t)$ denotes the local population size of each unique polymer species i on site x at time t, and N_{tot} is the total number of polymers on the local site x at time t (*i.e.* $N_{tot}(x,t) = \sum_i N_i(x,t)$). Local diversity measures the statistical diversity of each lattice site and is spatially averaged over all sites x to yield the average local diversity

$$\langle S_L(t)\rangle = \frac{\sum_x S_L(x,t)}{4096} \tag{26}$$

where 4096 is the total number of local populations (lattice sites) on our 64×64 square lattice. It is this spatially-averaged local diversity that is reported in our ensemble averaged data analysis (with spatial maps of local diversity included for simulation snapshots). Average local diversity therefore provides a statistical measure of the extent to which multiple sequences coexist on the same lattice site. As such, it provides a measure of diffusive mixing of populations and competition, whereby spatial regions with low local diversity result either from low diffusivity or high rates of local resource competition that result in fewer unique species.

A.3.2 Similarity Index

As an added measure of the spatial diversity of the polymer population we utilize a generalization of the of the Jaccard similarity index used in population ecology [75, 76]. This index provides a statistical measure of the similarity between next-nearest neighbor "communities" or lattice sites. Adopting this concept to chemical evolution, the similarity between two lattice sites α and β containing N_α and N_β unique sequences, respectively, and sharing $N_{\alpha\beta}$ mutual sequences is given by:

$$I_{\alpha\beta}(t) = \frac{R_\alpha(t)R_\beta(t)}{R_\alpha(t) + R_\beta(t) - R_\alpha(t)R_\beta(t)} \tag{27}$$

where R_α is the sum of the relative abundances of the shared species at site α:

$$R_\alpha(t) \equiv \sum_{i}^{N_{\alpha\beta}} \frac{N_i(\alpha,t)_{shared}}{N_{tot}(\alpha,t)} \ . \tag{28}$$

Here $N_i(\alpha,t)_{shared}$ is the abundance of the i-th species shared between sites α and β that are on site α at time t, and N_{tot} is the sum total of all polymers of both shared and unshared species at site α (*i.e.* $N_{tot}(\alpha,t) = \sum_i N_i(\alpha,t)$). An equivalent definitions for R_β is obtained with the substitution $\alpha \leftrightarrow \beta$.

A map of similarity of nearest-neighbor sites is produced by calculating the pairwise similarity index $I_{\alpha j}$ at the site α, where the index j runs over the eight nearest neighbor communities $j = 1, 2, ...8$. For each site, the indices $I_{\alpha j}$ are averaged to yield the site specific value:

$$I(x_\alpha, y_\alpha) = \frac{\sum_j I_{\alpha j}}{8} \qquad j = 1, 2, ...8 \tag{29}$$

The value $I(x_\alpha, y_\alpha)$ is recorded as the similarity index for site α in the similarity map. This procedure provides a method of identifying regions in the system occupied by highly similar populations of polymers as well as aiding in visual identification of clustered regions as shown in the Supporting Figures below.

A.3.3 Sequence Space Exploration Rate, Extant Species, and Species Lifetime

The total number of species introduced to the system, $N_S(t)$, is tracked over the entire course of each simulation and provides a measure of the total size of sequence space explored as a function of time. The rate of exploration of sequence space $R_S(t)$, *i.e.* the species production rate, is therefore calculated as the slope of the time-dependent total species, *e.g.* $R_S(t) = \dot{N}_S(t)$, discretized as $R_S(t_i) = \frac{N_S(t_i) - N_S(t_{i-1})}{2}$, which achieves a constant (within small fluctuations) and nonzero value during the quasi steady-state system evolution. $R_S(t)$ is time-averaged to obtain a mean exploration rate for each simulation run. Since the total number of extant species, N_{ES} (a directly measured value) achieves a constant value (within small fluctuations) during the quasi steady-state evolution, the number of species created per unit time is equal to the number destroyed as the extant population explores sequence space. An estimate of the mean species lifetime τ_S is therefore calculated as $\tau_S = N_{ES}/R_S$.

B Figures and Supporting Figures and Legends

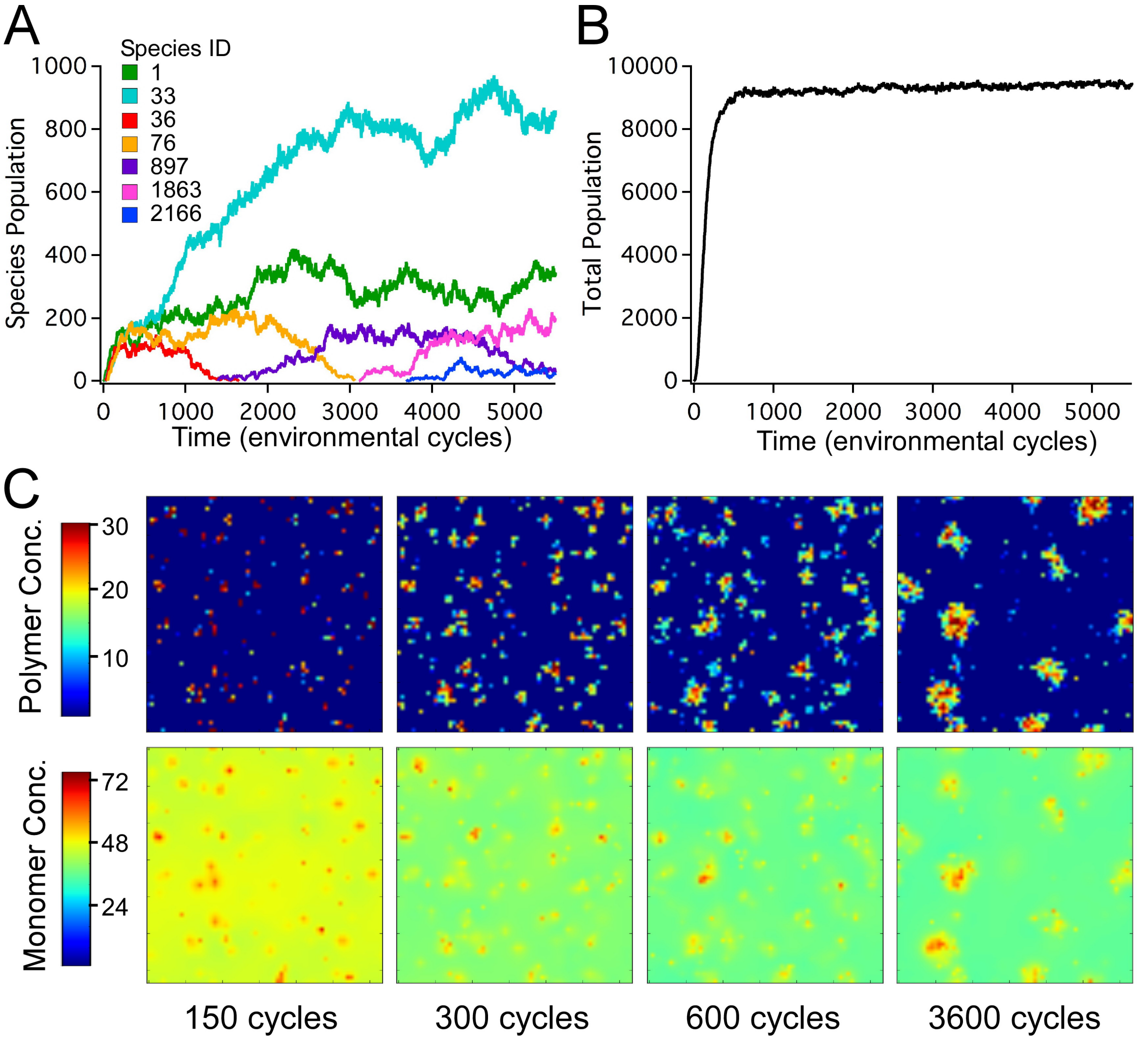

Figure 1. Sequence evolution of the polymer pool. A: Time evolution of the populations of seven specific sequences; B: Time evolution of the total polymer population; C: Spatial snapshots of the total polymer and monomer concentrations at four representative times. Species ID indicates the order of appearance of the first individual of a particular sequence in the polymer pool. The kinetic rate constants are $k_s = 10^{-7}$, $k_r = 10^{-4}$, and $k_h = 0.1$, and polymer and monomer diffusivities are set as $k_p = 0.001$ and $k_m = 10.0$ sites/cycle, respectively. Units of time are in number of cycles.

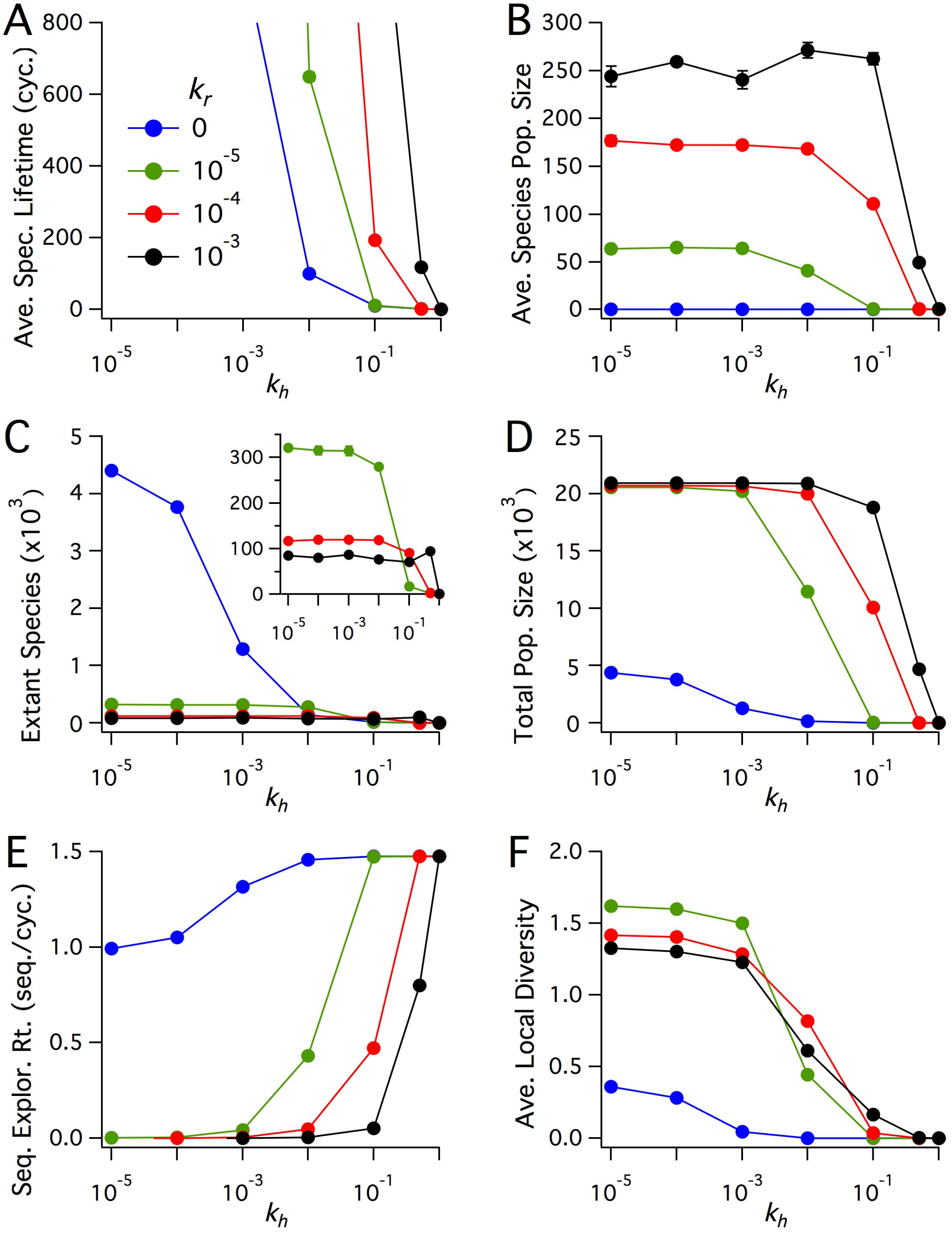

Figure 2. Exploring kinetic parameter space. Plots of quasi steady-state values of Average Species Lifetime (in units of number of cycles), Average Species Population, Extant Species, Total Population Size, Sequence Exploration Rate (in units of number of novel sequences generated per cycle), and Average Local Diversity. Time averages were taken from $t = 2500 - 5000$ cycles, and each point is the ensemble average over five realizations. Error bars denote the sample standard deviation (most are smaller than symbols). The rate constant for spontaneous sequence nucleation is $k_s = 10^{-7}$, and the diffusion rate constants are $k_p = 0.01$ sites/cycle and $k_m = 1.0$ sites/cycle. The blue data set shows the reference case with $k_r = 0$, where no polymers replicate.

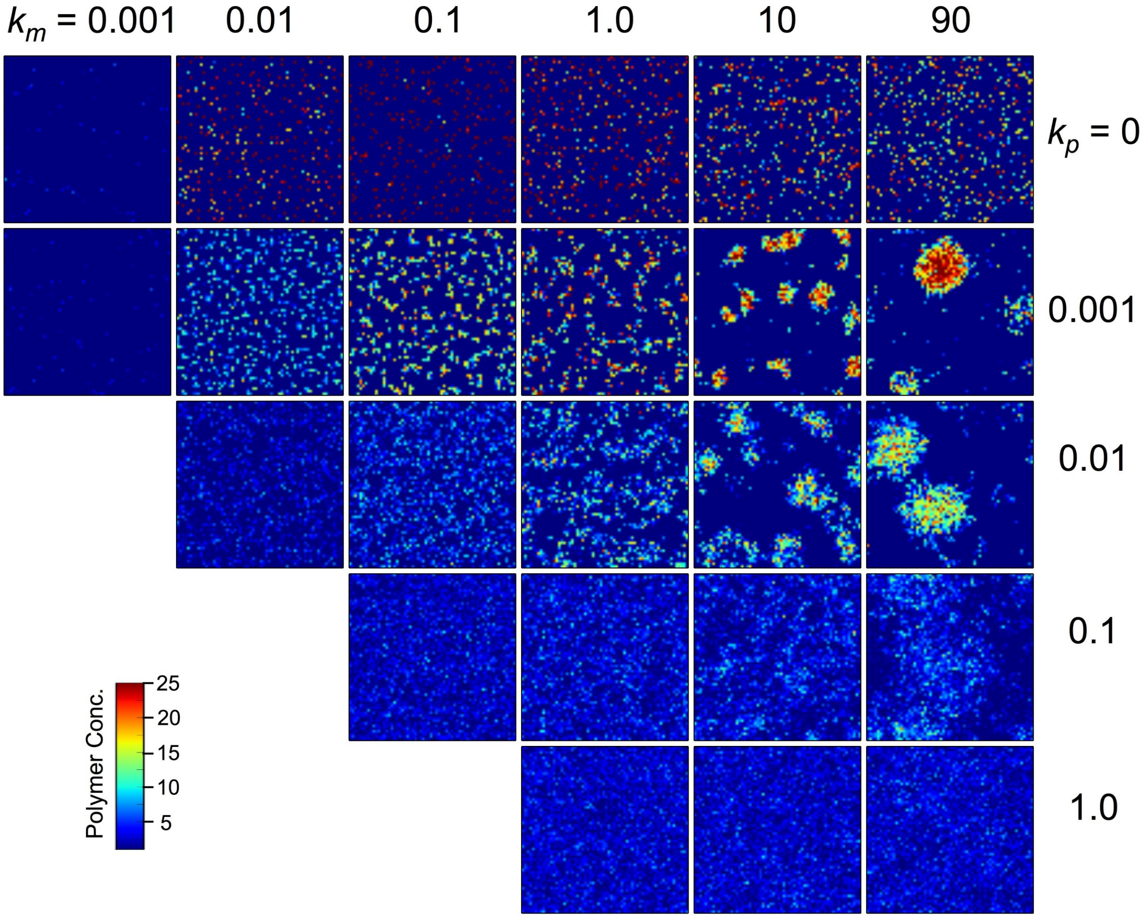

Figure 3. Spatial maps of polymer density. Each column of images corresponds to simulations run with a different value for monomer diffusivity k_m (in sites/cycle), and each row has a different value for polymer diffusivity k_p (in sites/cycle). All data shown are for kinetic rate constants $k_s = 10^{-7}$, $k_r = 10^{-4}$, and $k_h = 0.1$. All maps correspond to $t = 3000$ cycles. The color scale is in units of polymers/site. Simulations were only run for cases in which monomer diffusivity is greater than or equal to the polymer diffusivity.

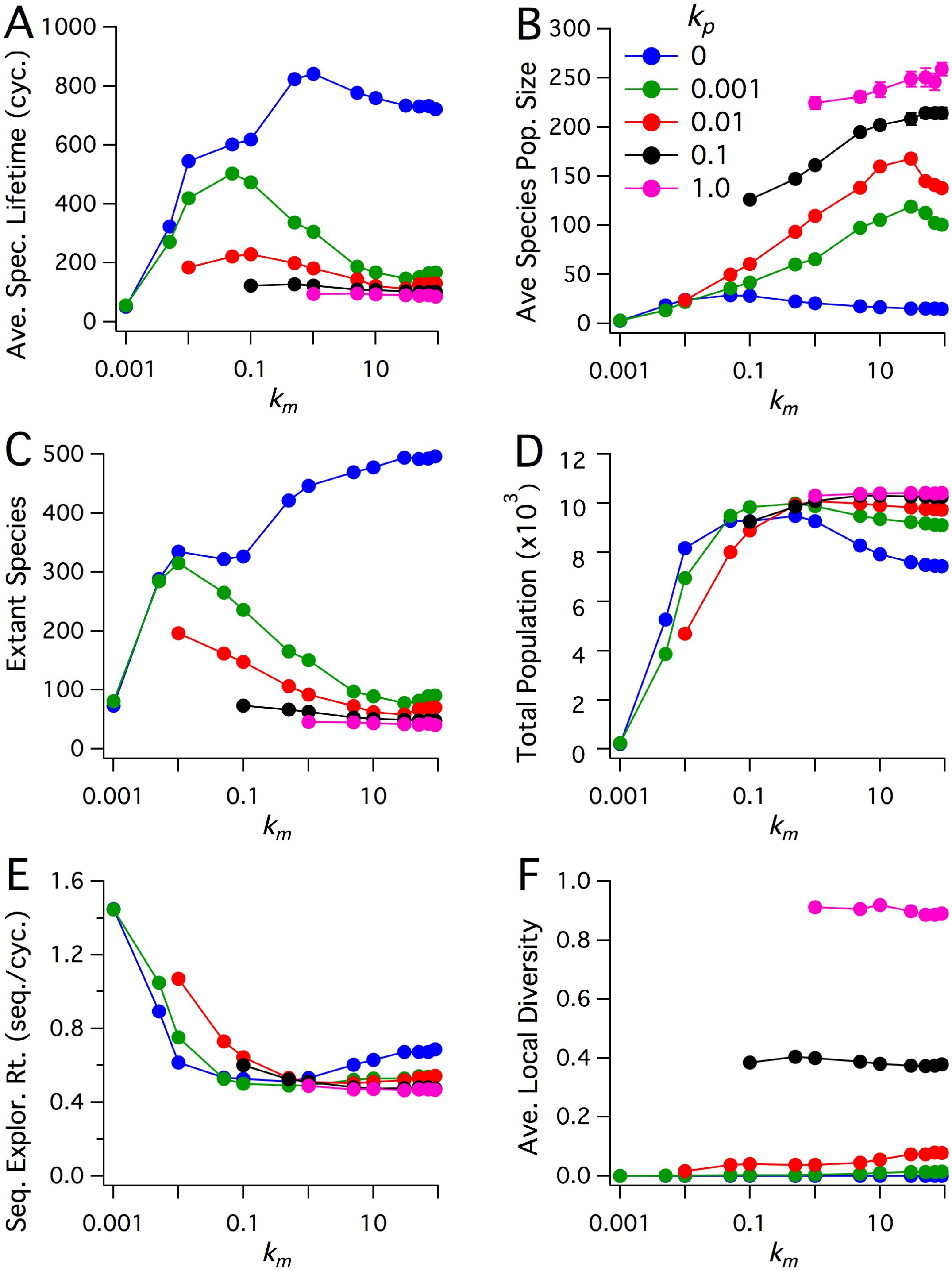

Figure 4. Exploring diffusive parameter space. Plots of quasi steady-state values of Average Species Lifetime (in units of number of cycles), Average Species Population, Extant Species, Total Population Size, Sequence Exploration Rate (in units of number of novel sequences generated per cycle), and Average Local Diversity. Time averages were taken from $t = 2500 - 5000$ cycles, and each point is the ensemble average over ten realizations. Error bars denote the sample standard deviation (most are smaller than symbols). The kinetic rate constants are $k_s = 10^{-7}$, $k_r = 10^{-4}$, and $k_h = 0.1$. The blue data set shows the case where the $k_p = 0$, where polymers are completely immobile.

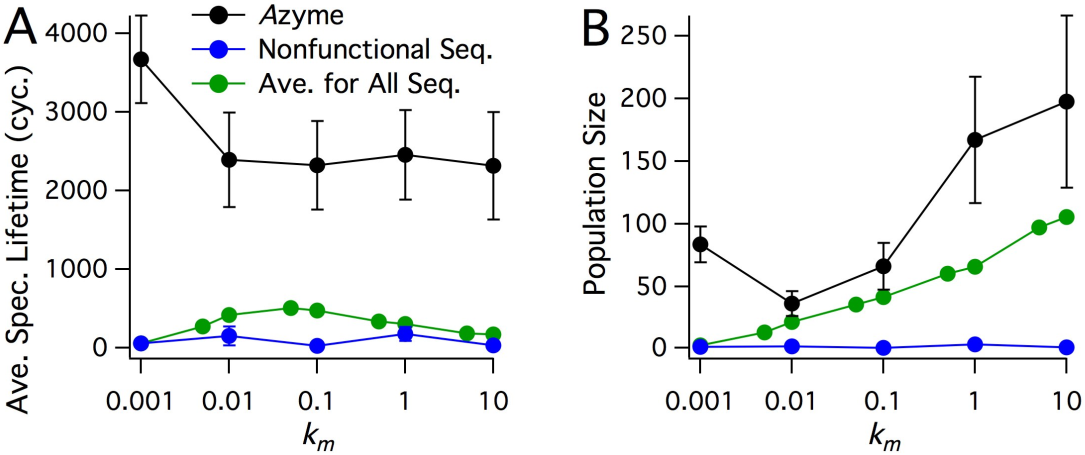

Figure 5. Selection for functional sequences. Plots illustrating the propagation of a functional *Azyme*, compared to nonfunctional sequences. A: Average Species Lifetime (in units of number of cycles), and B: Average Population Size. Each data point is the ensemble average over twenty-five runs, with error bars denoting the sample standard deviation. Kinetic rate constants are $k_s = 10^{-7}$, $k_r = 10^{-4}$, and $k_h = 0.1$, with a polymer diffusion rate constant of $k_p = 0.001$ sites/cycle. The green points represent overall population statistics for realizations with no *Azyme* (plotted in green in Figure 4). The black points represent statistics for a single functional *Azyme*. The blue points represent statistics for a nonfunctional sequence introduced at the same location and cycle as in the *Azyme* simulations, except with no functionality.

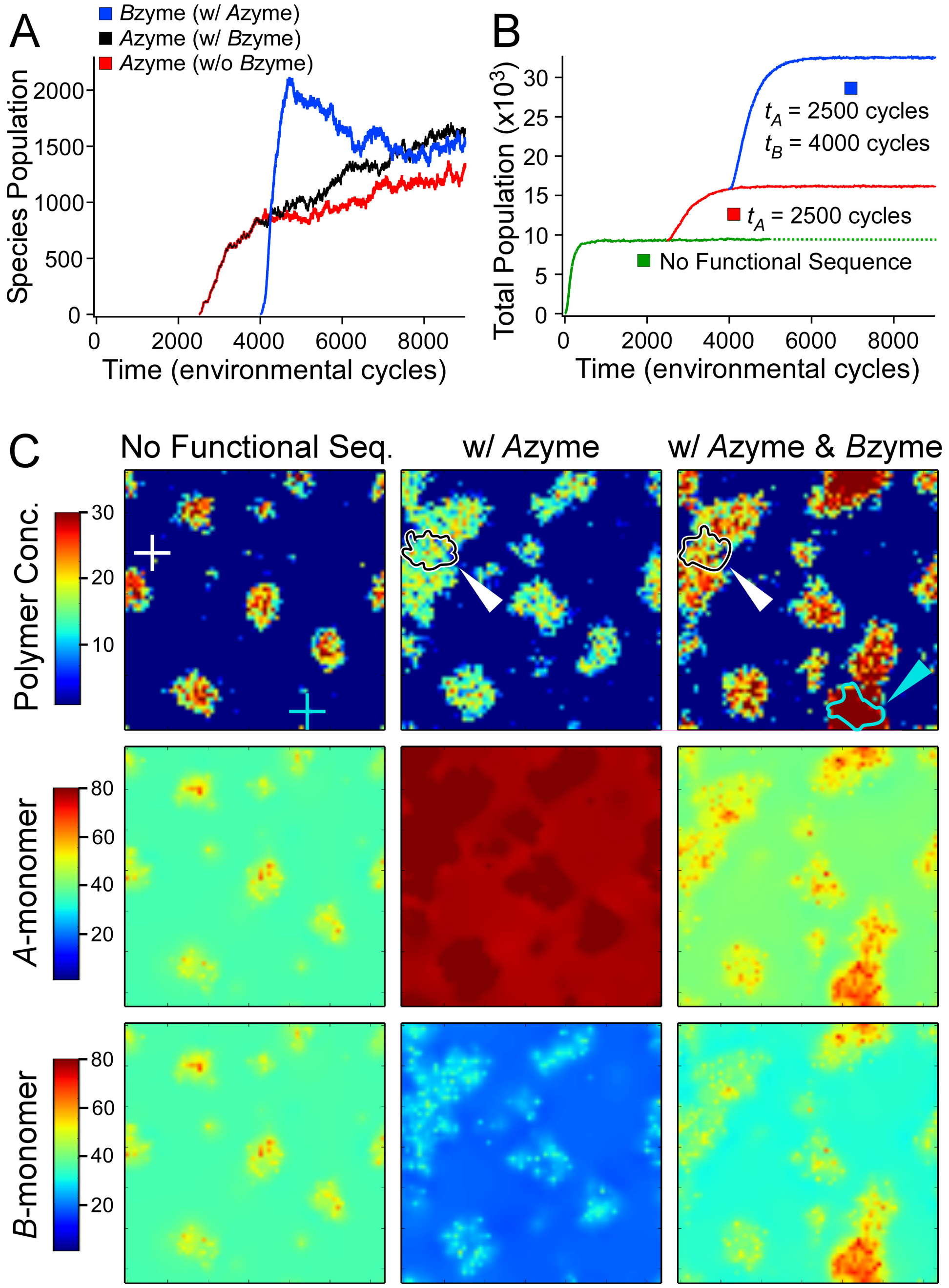

Figure 6. Spatial distribution maps for no functional species, one functional species, and two functional species. The three scenarios shown are all identical up to $t = 2500$ cycles, at which time the system has achieved a quasi-steady state distribution. In the first scenario, no functional sequences appear. In the second scenario, a functional $Azyme$ appears at $t_A = 2500$. In the third scenario, the same functional $Azyme$ appears at $t_A = 2500$, and a functional $Bzyme$ also appears at $t_B = 4000$. A: Time evolution of the Species Populations of the $Azyme$ and $Bzyme$. The units of time are in number of cycles. The red curve corresponds to the second scenario, having only the $Azyme$, while the black and blue curves correspond to the third scenario with both enzymes emerging. B: The time evolution of the Total Polymer Population for the three scenarios. C: The spatial distribution of the polymer (total) and monomer concentrations, at $t = 5000$ cycles. White arrow indicates contour containing 95% of $Azyme$ polymers, cyan arrow indicates contour containing 95% of $Bzyme$ polymers. Kinetic rate constants are $k_s = 10^{-7}$, $k_r = 10^{-4}$, and $k_h = 0.1$, and diffusive rate constants are $k_p = 0.01$ and $k_m = 1.0$ sites/cycle.

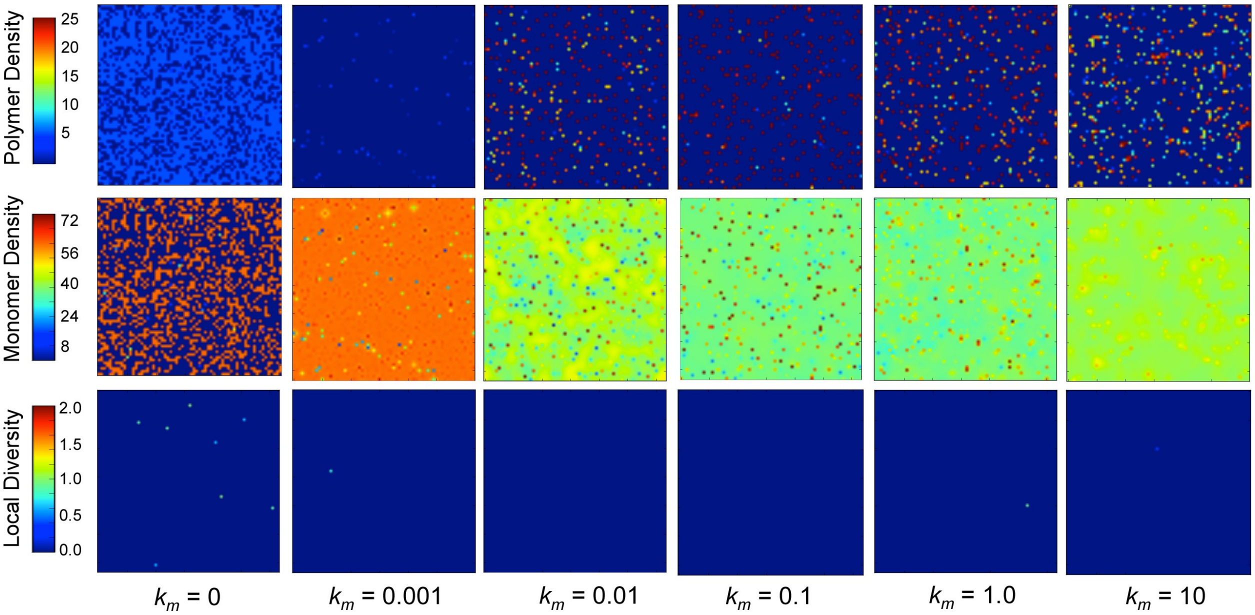

Figure 7. Figure S1. Spatial maps for $k_p = 0$ sites/cycle. Spatial maps of polymer density (top row), $A = B$ monomer density (middle), and local diversity (bottom) for $k_p = 0$ sites/cycle, with monomer diffusivity increasing from left to right. Polymers do not diffuse and as such are indefinitely stuck on their nucleation site. Snapshots are taken at $t = 2000$ cycles. The kinetic rate constants are $k_s = 10^{-7}$, $k_r = 10^{-4}$, and $k_h = 0.1$.

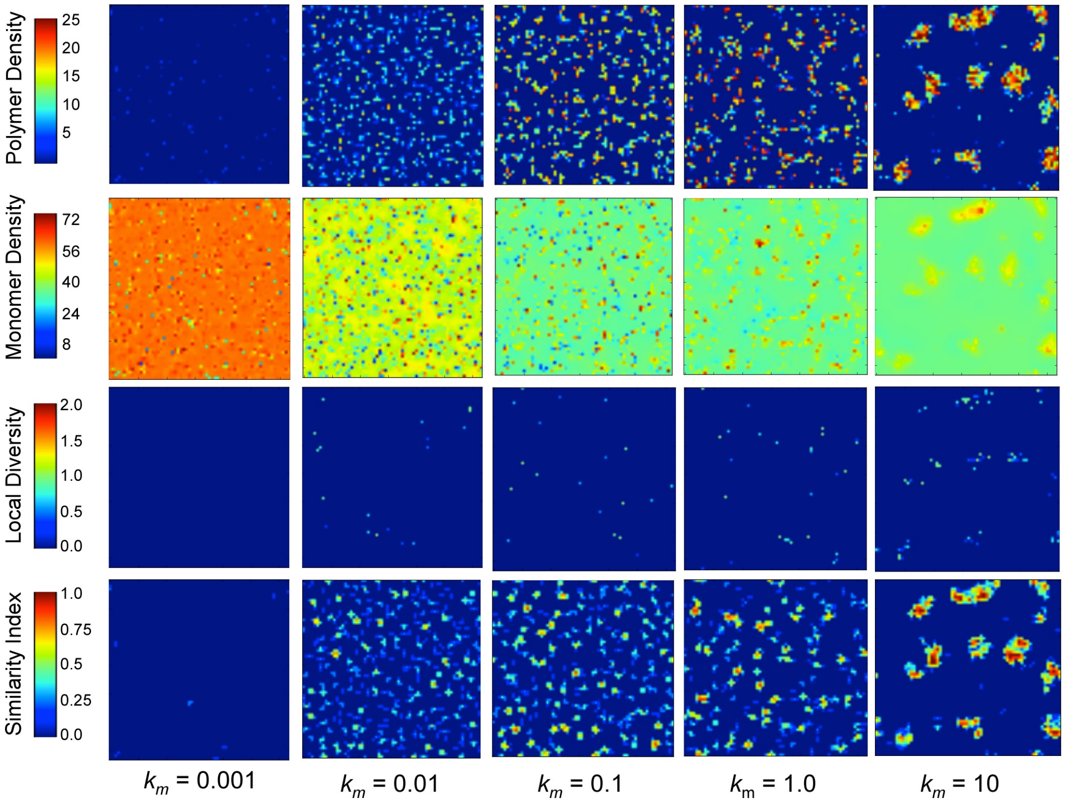

Figure 8. Figure S2. Spatial maps for $k_p = 0.001$ sites/cycle. Spatial maps of polymer density (top row), $A = B$ monomer density (second row), local diversity (third row), and similarity index (bottom row) for $k_p = 0.001$ sites/cycle, with monomer diffusivity increasing from left to right. Snapshots are taken at $t = 2000$ cycles. The kinetic rate constants are $k_s = 10^{-7}$, $k_r = 10^{-4}$, and $k_h = 0.1$.

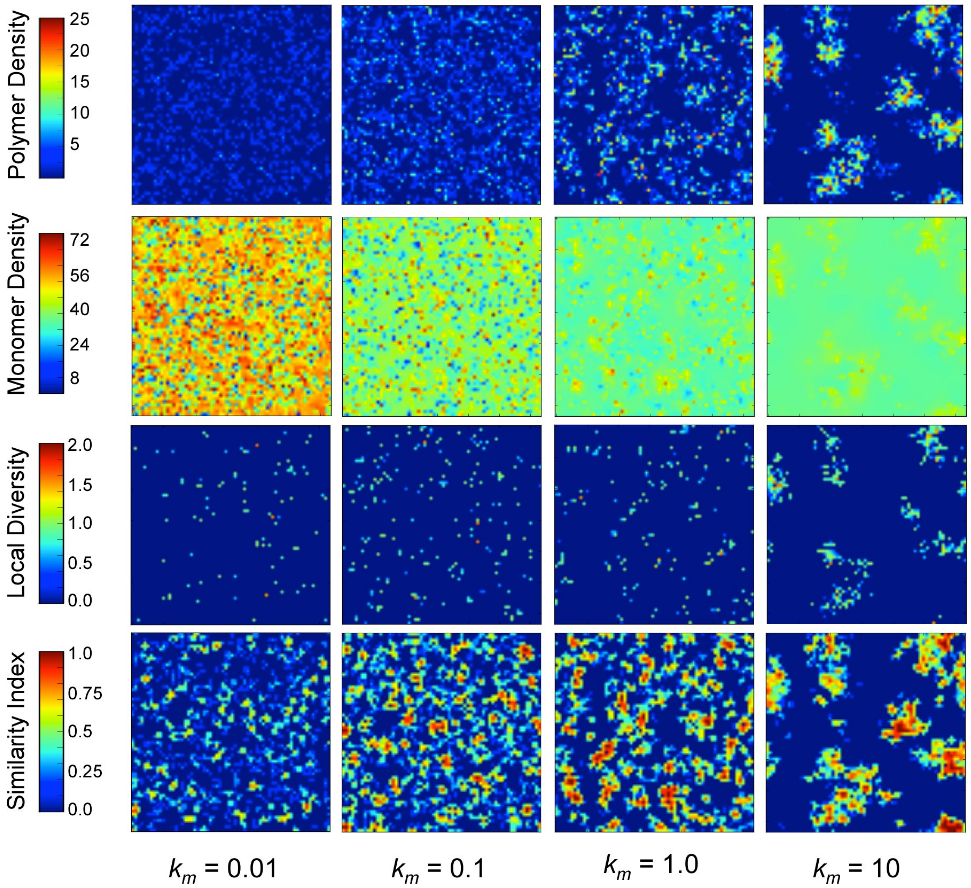

Figure 9. Figure S3. Spatial maps for $k_p = 0.01$ sites/cycle. Spatial maps of polymer density (top row), $A = B$ monomer density (second row), local diversity (third row), and similarity index (bottom row) for $k_p = 0.01$ sites/cycle, with monomer diffusivity increasing from left to right. Snapshots are taken at $t = 2000$ cycles. The kinetic rate constants are $k_s = 10^{-7}$, $k_r = 10^{-4}$, and $k_h = 0.1$.

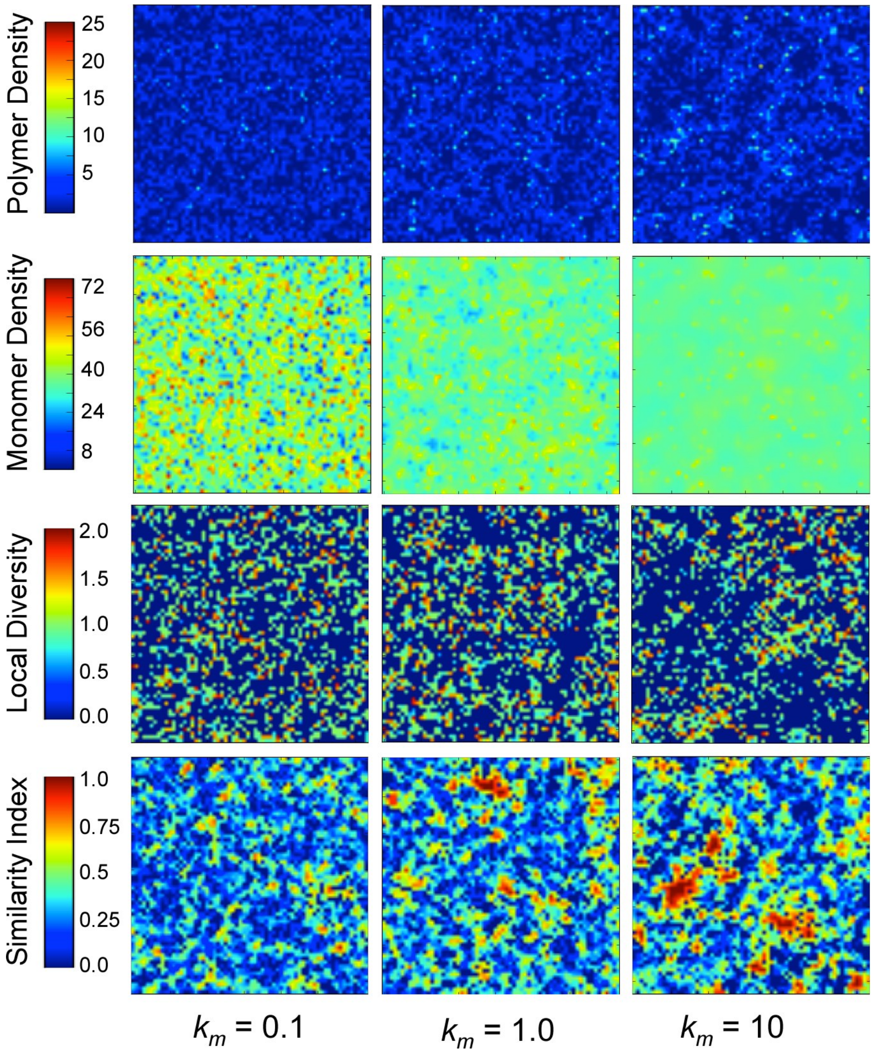

Figure 10. Figure S4. Spatial maps for $k_p = 0.1$ sites/cycle. Spatial maps of polymer density (top row), $A = B$ monomer density (second row), local diversity (third row), and similarity index (bottom row) for $k_p = 0.1$ sites/cycle, with monomer diffusivity increasing from left to right. Snapshots are taken at $t = 2000$ cycles. The kinetic rate constants are $k_s = 10^{-7}$, $k_r = 10^{-4}$, and $k_h = 0.1$.

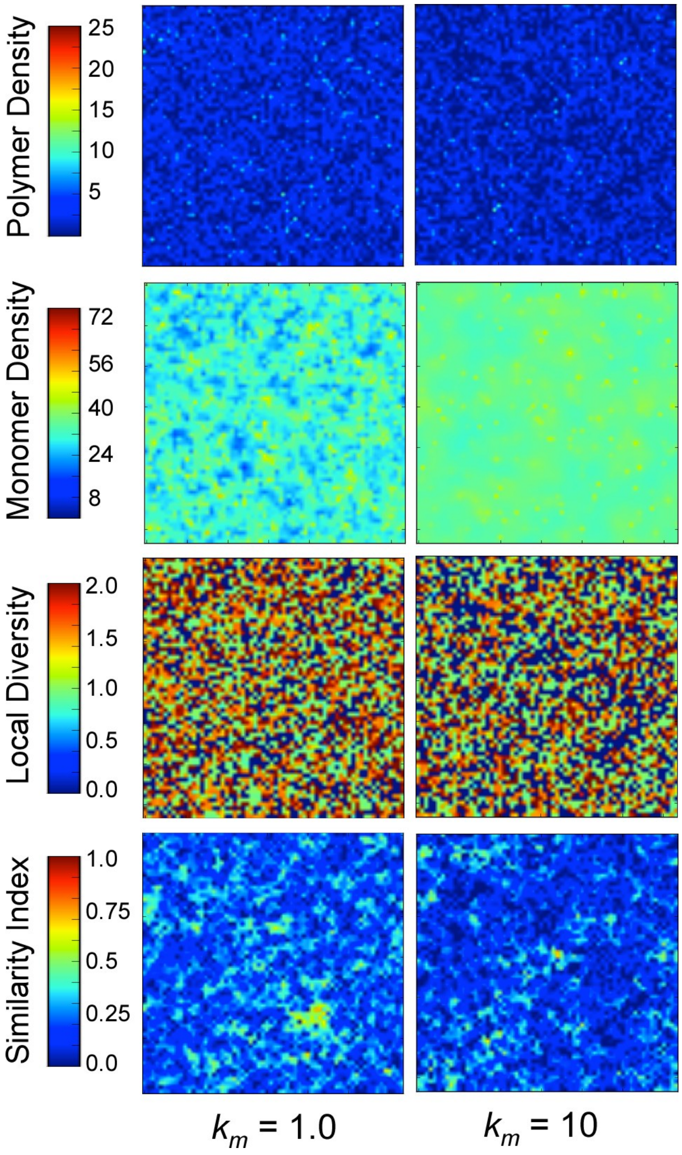

Figure 11. Figure S5. Spatial maps for $k_p = 1.0$ sites/cycle. Spatial maps of polymer density (top), $A = B$ monomer density (second row), local diversity (third row), and similarity index (bottom row) for $k_p = 1.0$ sites/cycle, with monomer diffusivity increasing from left to right. Snapshots are taken at $t = 2000$ cycles. The kinetic rate constants are $k_s = 10^{-7}$, $k_r = 10^{-4}$, and $k_h = 0.1$.

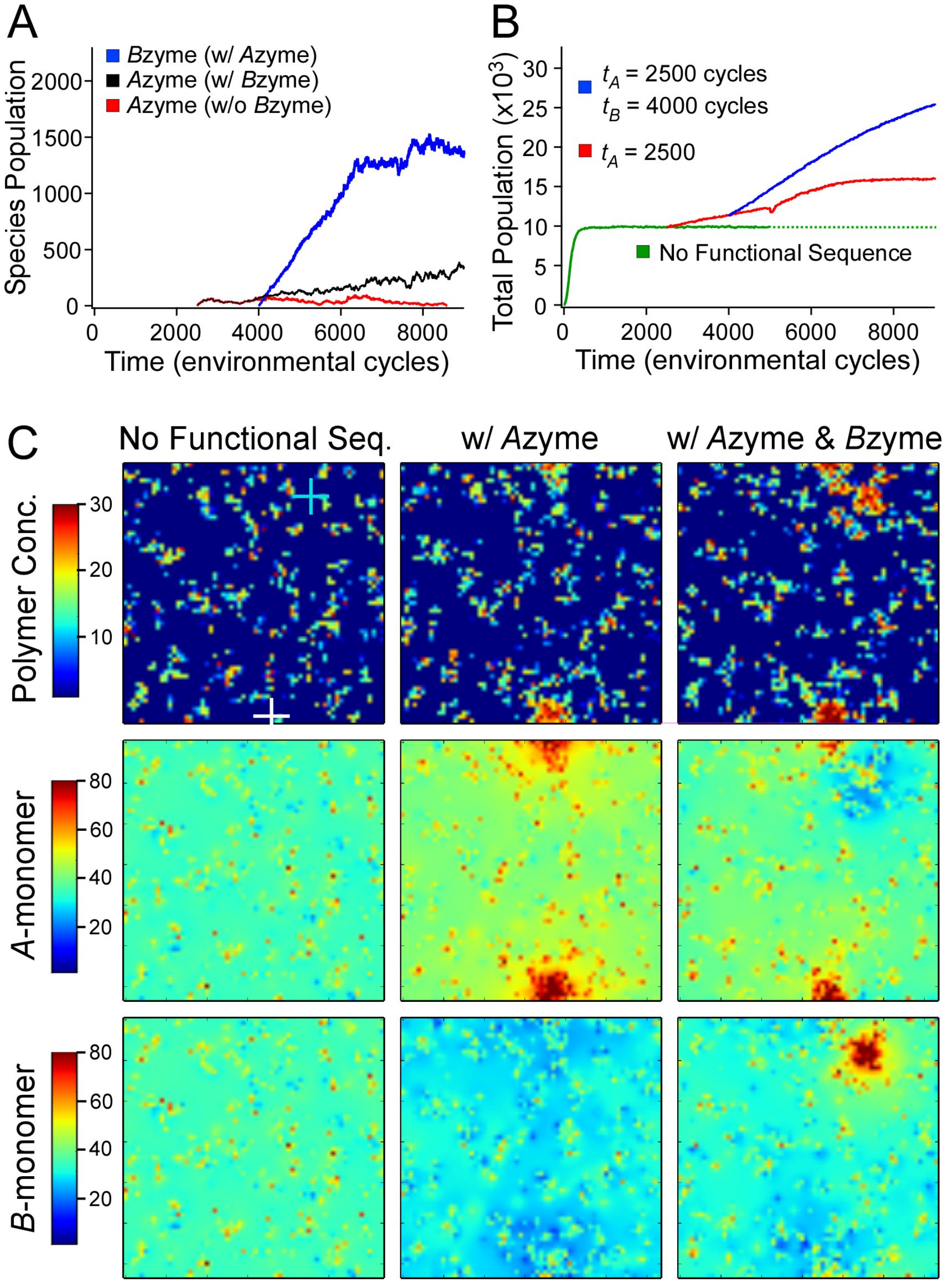

Figure 12. Figure S6. Spatial distribution maps for no functional species, one functional species, and two functional species. The three scenarios shown are all identical up to $t = 2500$ cycles, at which time the system has achieved a quasi-steady state distribution. In the first scenario, no functional sequences appear. In the second scenario, a functional Azyme appears at $t_A = 2500$. In the third scenario, the same functional Azyme appears at $t_A = 2500$ cycles, and the functional Bzyme appears at $t_B = 4000$ cycles. In Panel A, the time evolution of the Species Populations of the Azyme and Bzyme is shown. The red curve corresponds to the second scenario, having only the Azyme, while the black and blue curves correspond to the third scenario with both enzymes emerging. Panel B shows the time evolution of the Total Polymer Population for the three scenarios. Panel C illustrates the spatial distribution of the polymer (total) and monomer concentrations, at $t = 5000$ cycles. Kinetic rates are $k_s = 10^{-7}$, $k_r = 10^{-4}$, and $k_h = 0.1$, and diffusive rates of $k_p = 0.001$ and $k_m = 1.0$ sites/cycle.